BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Attorney Docket No. 826.1913

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Yasuko SOGA et al.

Application No.: Not Yet Assigned | Group Art Unit: Not Yet Assigned

Filed: December 17, 2003 | Examiner: Not Yet Assigned

For: VIDEO TEXT PROCESSING APPARATUS

SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN APPLICATION IN ACCORDANCE WITH THE REQUIREMENTS OF 37 C.F.R. § 1.55

Commissioner for Patents
PO Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s) herewith a certified copy of the following foreign application:

Japanese Patent Application No(s). 2002-378577

Filed: December 26, 2002

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing date(s) as evidenced by the certified papers attached hereto, in accordance with the requirements of 35 U.S.C. § 119.

Respectfully submitted,

STAAS & HALSEY LLP

Date: December 17, 2003 By: _____

J. Randall Beckers
Registration No. 30,358

1201 New York Ave, N.W., Suite 700
Washington, D.C. 20005
Telephone: (202) 434-1500
Facsimile: (202) 434-1501

JAPAN PATENT OFFICE

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: December 26, 2002

Application Number: Patent Application No. 2002-378577
[ST.10/C]: [JP2002-378577]

Applicant(s): FUJITSU LIMITED

October 31, 2003

Commissioner,
Japan Patent Office Yasuo IMAI

Certificate No. P2003-3090803

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００２年１２月２６日

出願番号
Application Number: 特願２００２－３７８５７７

[ST. 10/C]: [ＪＰ２００２－３７８５７７]

出願人
Applicant(s): 富士通株式会社

２００３年１０月３１日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号　出証特２００３－３０９０８０３

【書類名】　特許願

【整理番号】　0252606

【特記事項】　特許法第３６条の２第１項の規定による特許出願

【提出日】　平成14年12月26日

【あて先】　特許庁長官殿

【国際特許分類】　G06K　9/00

G06F 16/00

【発明の名称】　Ｖｉｄｅｏ　Ｔｅｘｔ　Ｐｒｏｃｅｓｓｉｎｇ　Ａｐｐａｒａｔｕｓ

【請求項の数】　8

【発明者】

【住所又は居所】　中国北京市朝陽区霄雲路２６号　鵬潤大厦Ｂ－１００３　富士通研究開発中心有限公司内

【氏名】　孫　俊

【発明者】

【住所又は居所】　神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

【氏名】　勝山　裕

【発明者】

【住所又は居所】　神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

【氏名】　直井　聡

【特許出願人】

【識別番号】　000005223

【氏名又は名称】　富士通株式会社

【代理人】

【識別番号】　100074099

【住所又は居所】　東京都千代田区二番町８番地２０　二番町ビル３Ｆ

【弁理士】

【氏名又は名称】　大菅　義之

【電話番号】　03-3238-0031

【選任した代理人】

【識別番号】　100067987

【住所又は居所】　神奈川県横浜市鶴見区北寺尾７－２５－２８－５０３

【弁理士】

【氏名又は名称】　久木元　彰

【電話番号】　045-573-3683

【手数料の表示】

【予納台帳番号】　012542

【納付金額】　35,000円

【提出物件の目録】

【物件名】　外国語明細書　1

【物件名】　外国語図面　1

【物件名】　外国語要約書　1

【包括委任状番号】　9705047

【プルーフの要否】　要

【書類名】　外国語明細書

【発明の名称】　Video Text Processing Apparatus

【特許請求の範囲】

1. A text change frame detection apparatus that selects a plurality of video frames including text contents from given video frames, characterized in that said apparatus comprises:

first frame removing means for removing redundant video frames from the given video frames;

second frame removing means for removing video frames that do not contain a text area from the given video frames;

third frame removing means for detecting and removing redundant video frames caused by image shifting from the given video frames; and

output means for outputting remaining video frames as candidate text change frames.

2. A text extraction apparatus that extracts at least one text line region from a given image, characterized in that said apparatus comprises:

edge image generation means for generating edge information of the given image;

stroke image generation means for generating a binary image of candidate character strokes in the given image by using the edge information;

stroke filtering means for removing a false stroke from the binary image by using the edge information;

text line region formation means for combining a plurality of strokes into a text line region;

text line verification means for removing a false character stroke from the text line region and reforming the text line region;

text line binarization means for binarizing the text line region by using a height of the text line region; and

output means for outputting a binary image of the text line region.

3. A program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

removing redundant video frames from the given video frames;

removing video frames that do not contain a text area from the given video frames;

detecting and removing redundant video frames caused by image shifting from the given video frames; and

outputting remaining video frames as candidate text change frames.

4. A program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

determining whether two image blocks in the same position in two video frames of given video frames are a valid block pair that has an ability to show a change of image contents;

calculating a similarity of two image blocks of the valid block pair and determining whether the two image

blocks are similar;

determining whether the two video frames are similar by using a ratio of a number of similar image blocks to a total number of valid block pairs; and

outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

5. A program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

generating a first binary image of a video frame of the given video frames;

determining a position of a text line region by using a horizontal projection and a vertical projection of the first binary image;

generating a second binary image of every text line region;

determining validity of a text line region by using a difference between the first binary image and the second binary image and a fill rate of a number of foreground pixels in the text line region to a total number of pixels in the text line region;

confirming whether a set of continuous video frames are non-text frames that do not contain a text area by using a number of valid text line regions in the set of continuous video frames; and

outputting remaining video frames after the non-text frames are removed, as candidate text change frames.

6. A program for a computer that selects a plurality of

video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

generating binary images of two video frames of the given video frames;

determining a vertical position of every text line region by using horizontal projections of the binary images of the two video frames;

determining a vertical offset of image shifting between the two video frames and a similarity of the two video frames in a vertical direction by using correlation between the horizontal projections;

determining a horizontal offset of the image shifting and a similarity of the two video frames in a horizontal direction by using correlation between vertical projections of every text line in the binary images of the two video frames; and

outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

7. A program for a computer that extracts at least one text line region from a given image, characterized in that the program directs the computer to perform a process comprising:

generating edge information of the given image;

generating a binary image of candidate character strokes in the given image by using the edge information;

removing a false stroke from the binary image by using the edge information;

combining a plurality of strokes into a text line

region;

removing a false character stroke from the text line region and reforming the text line region;

binarizing the text line region by using a height of the text line region; and

outputting a binary image of the text line region.

8. A program for a computer that extracts at least one text line region from a given image, characterized in that the program directs the computer to perform a process comprising:

generating an edge image of the given image;

generating a binary image of candidate character strokes in the given image by using the edge image;

checking an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels indicating an edge in the edge image;

determining that the stroke is a valid stroke if the overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold;

removing the invalid stroke; and

outputting information of remaining strokes in the binary image of the candidate character strokes.

【発明の詳細な説明】

Technical Field to which Invention Belongs

The present invention relates to a video image processing apparatus, more specifically to a text image extraction apparatus for e-Learning video. The text change

frame detection apparatus locates the video frames that contain text information. The text extraction apparatus extracts the text information out of the video frames and send the extracted text information to an optical character recognition (OCR) engine for recognition.

Prior Art

Text retrieval in video and image is a very important technique and has a variety of application, such as storage capacity reduction, video and image indexing, and digital library, etc.

The present invention focuses on a special type of video - e-Learning video, which often contains a large amount of text information. In order to efficiently retrieve the text content in the video, two techniques are needed: text change frame detection in video and text extraction from image. A text change frame is a frame that marks the change of text content in a video. The first technique fast browses the video and selects those video frames that contain text area. The second technique then extracts the text information from those video frames and sends them to an OCR engine for recognition.

Text change frame detection technique can be regarded as a special case of scene change frame detection technique. The techniques for detecting the scene change frame that marks the changes of the content in video from a plurality of frames in a video have been studied actively in recent years. Some methods focus on the intensity difference between frames, some methods focus on

the difference of color histogram and the texture. However, these methods are not suitable for text change frame detection in video, especially in e-Learning field.

Take presentation video - a typical e-Learning video as example, in which the video frame often contains a slide image. Examples of slide image include the PowerPoint® image and the film image from a projector. The change of the content of slide will not cause a dramatic change in color and texture. Also, the focus of the video camera often moves around in a slide image during the talk, which causes image shifting. Image shifting also occurs when the speaker moving his or her slides. These content shifting frames will be marked as scene change frames by conventional methods. Another drawback of the conventional method is that they can not tell directly whether a frame contains text information.

Another way to extract text change frame from video is performing text extraction method on every frame in the video and judging whether the content has been changed. The problem of such strategy is that it is very time consuming.

After the text change frames are detected, a text extraction method should be used to extract the text lines from the frames. Many methods are proposed to extract the text lines from video and static image (see non-patent documents 1 and 2, for example).

Also, some patents related to this field have been published (see patent documents 1, 2 and 3, for example).

These methods will meet problems when deal with

video frame in e-Learning. The characters in e-Learning video image always have very small size, also the boundaries of these characters are very dim, and there are many disturbances around the text area, such like the bounding box of text line, the shading and occlusion of human body, etc.

Non-Patent Document 1

V. Wu, R. Manmatha, and E. M. Riseman, "TextFinder: An Automatic System to Detect and Recognize Text in Images," IEEE transactions on Pattern Analysis and Machine Intelligence, VOL. 21, NO. 11, pp. 1224-1229, November, 1999.

Non-Patent Document 2

T. Sato, T. Kanade, E. Hughes, M. Smith, and S. Satoh, "Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Captions," ACM Multimedia Systems Special Issue on Video Libraries, February, 1998.

Patent Document 1

U.S. Patent No. 6,366,699, specification.

Patent Document 2

U.S. Patent No. 5,465,304, specification.

Patent Document 3

U.S. Patent No. 5,307,422, specification.

Problems to be Solved by Invention

However, there are the following problems in the above mentioned conventional video image processing.

It is very time consuming to perform text extraction method on every frame in the video and judge whether the content has been changed.

The characters in e-Learning video image always have very small size, also the boundaries of these characters are very dim, and there are many disturbances around the text area. Therefore, the conventional text extraction method will leave many false character strokes in the final binary image, which give a wrong recognition result in the following OCR stage.

It is an object of the present invention to select the candidate text change frames from a plurality of video frames in a fast speed, while keeping a high recall rate, which is defined as the rate of the number of extracted correct text change frames to the total number of correct text change frames.

It is another object of the present invention to provide a scheme for efficiently detecting the text region in the text change frame, removing as much as possible the false character strokes, and providing a binarized image for every text line.

Means for Solving Problems

The above objects are fulfilled by a video text processing apparatus for fast selecting from all frames in a video those frames that contain text contents, marking

the region of each text line in the text frame and outputting the text line in a binary form, comprising a text change frame detection apparatus for fast selecting text frames in the video and a text extraction apparatus for extracting the text lines in the text frame. The binary form is, for example, represented by black pixels corresponding to background and white pixels corresponding to character strokes.

The first text change frame detection apparatus comprises first frame removing means, second frame removing means, third frame removing means and output means, and selects a plurality of video frames including text contents from given video frames. The first frame removing means removes redundant video frames from the given video frames. The second frame removing means removes video frames that do not contain a text area from the given video frames. The third frame removing means detects and removes redundant video frames caused by image shifting from the given video frames. The output means outputs remaining video frames as candidate text change frames.

The second text change frame detection apparatus comprises image block validation means, image block similarity measurement means, frame similarity judgment means and output means, and selects a plurality of video frames including text contents from given video frames.

The image block validation means determines whether two image blocks in the same position in two video frames of given video frames are a valid block pair that has an

ability to show a change of image contents. The image block similarity measurement means calculates a similarity of two image blocks of the valid block pair and determines whether the two image blocks are similar. The frame similarity judgment means determines whether the two video frames are similar by using a ratio of the number of similar image blocks to the total number of valid block pairs. The output means outputs remaining video frames after a similar video frame is removed, as candidate text change frames.

The third text change frame detection apparatus comprises fast and simple image binarization means, text line region determination means, rebinarization means, text line confirmation means, text frame verification means and output means, and selects a plurality of video frames including text contents from given video frames. The fast and simple image binarization means generates a first binary image of a video frame of the given video frames. The text line region determination means determines a position of a text line region by using a horizontal projection and a vertical projection of the first binary image. The rebinarization means generates a second binary image of every text line region. The text line confirmation means determines validity of a text line region by using a difference between the first binary image and the second binary image and a fill rate of the number of foreground pixels in the text line region to the total number of pixels in the text line region. The text frame verification means confirms whether a set of

continuous video frames are non-text frames that do not contain a text area by using the number of valid text line regions in the set of continuous video frames. The output means outputs remaining video frames after the non-text frames are removed, as candidate text change frames.

The fourth text change frame detection apparatus comprises fast and simple image binarization means, text line vertical position determination means, vertical shifting detection means, horizontal shifting detection means and output means, and selects a plurality of video frames including text contents from given video frames. The fast and simple image binarization means generates binary images of two video frames of the given video frames. The text line vertical position determination means determines a vertical position of every text line region by using horizontal projections of the binary images of the two video frames. The vertical shifting detection means determines a vertical offset of image shifting between the two video frames and a similarity of the two video frames in a vertical direction by using correlation between the horizontal projections. The horizontal shifting detection means determines a horizontal offset of the image shifting and a similarity of the two video frames in a horizontal direction by using correlation between vertical projections of every text line in the binary images of the two video frames. The output means outputs remaining video frames after a similar video frame is removed, as candidate text change frames.

After the candidate text change frames in the videc are detected by the text change frame detection apparatus, the image of every frame is then sent to the text extraction apparatus for text extraction.

The first text extraction apparatus comprises edge image generation means, stroke image generation means, stroke filtering means, text line region formation means, text line verification means, text line binarization means and output means, and extracts at least one text line region from a given image. The edge image generation means generates edge information of the given image. The stroke image generation means generates a binary image of candidate character strokes in the given image by using the edge information. The stroke filtering means removes the false strokes from the binary image by using the edge information. The text line region formation means combines a plurality of strokes into a text line region. The text line verification means removes a false character stroke from the text line region and reforms the text line region. The text line binarization means binarizes the text line region by using a height of the text line region. The output means outputs a binary image of the text line region.

The second text extraction apparatus comprises edge image generation means, stroke image generation means, stroke filtering means and output means, and extracts at least one text line region from a given image. The edge image generation means generates an edge image of the given image. The stroke image generation means generating

a binary image of candidate character strokes in the given image by using the edge image. The stroke filtering means checks an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels indicating an edge in the edge image, determines that the stroke is a valid stroke if the overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold, and removes the invalid stroke. The output means outputs information of remaining strokes in the binary image of the candidate character strokes.

After the text line regions are extracted by the text extraction apparatus, they are sent to an OCR engine for recognition.

Embodiments of Invention

The embodiments of the present invention are described below in detail by referring to the drawings.

Fig. 1 shows the configuration of the video text processing apparatus according to the present invention. The input of the apparatus is an existing video data 101 or living video stream from a television (TV) video camera 102, the input video data is first decomposed into continuous frames by a video decomposition unit 103. Then a text change frame detection apparatus 104 is used to find the candidate text change frames in the video frames. The text change frame detection apparatus will greatly reduce the total processing time. After that, a text extraction apparatus 105 is enforced on every candidate

text change frame to detect text lines (text areas) in the frames and output the images of the text lines to a database 106 for further OCR processing.

Fig. 2 shows the processing flow chart of the video text processing apparatus shown in Fig. 1. A process in S201 is performed by the video decomposition unit 103, processes in S202 to S204 are performed by the text change frame detection apparatus 104, and processes in S205 to S210 are performed by the text extraction apparatus 105.

First the input video is decomposed into continuous frames (S201). Then frame similarity measurement is performed to measure the similarity of two nearby frames (S202). If the two frames are similar, then the second frame is removed. Next text frame detection and verification is performed to judge whether the remaining frames from the process in S202 contain text lines (S203). If a frame does not contain a text line, the frame is removed. Image shifting detection is further performed to determine whether image shifting exists in two frames (S204). If so, the second frame is removed. The output of the text change frame detection apparatus 104 is a group of candidate text change frames.

For every candidate text change frame, edge image generation is performed to generate the edge image of the frame (S205). Then stroke generation is performed to generate the stroke image based on edge information (S206). Next stroke filtering is performed to remove false strokes based on edge information (S207). Text line region formation is further performed to connect individual

strokes into a text line (S208). After that, text line verification is performed to remove false strokes in a text line and re-form the text line (S209). Finally, text line binarization is performed to produce the final binary image of the text line (S210). The final output is a serial of binary text line images that will be processed by an OCR engine for recognition.

Fig. 3 shows the configuration of the text change frame detection apparatus 104 shown in Fig. 1. The input video frames are first sent to a frame similarity measurement unit 301 for deleting duplicate frames, then a text frame detection and verification unit 302 is used to check whether a frame contains text information. Next, an image shifting detection unit 303 is used to remove redundant frames that caused by image shifting. The frame similarity measurement unit 301, the text frame detection and verification unit 302 and the image shifting detection unit 303 correspond to the first, second and third frame removing means, respectively. The text change frame detection apparatus 104 is very suitable to detect text change frame in e-Learning video. It can remove duplicate video frames, shifting video frames as well as video frames that do not contain text area in a very fast speed while keeping a high recall rate.

Fig. 4 shows the configuration of the frame similarity measurement unit 301 shown in Fig. 3. The frame similarity measurement unit 301 includes an image block validation unit 311, an image block similarity measurement unit 312, and a frame similarity judgment unit 313. The

image block validation unit 311 determines whether two image blocks in the same position in two video frames are a valid block pair. A valid block pair is an image block pair that has the ability to show the change of the image content. The image block similarity measurement unit 312 calculates the similarity of two image blocks of the valid block pair and determines whether the two image blocks are similar. The frame similarity judgment unit 313 determines whether the two video frames are similar by using a ratio of the number of similar image blocks to the total number of valid block pairs. According to the frame similarity measurement unit 301, duplicate frames are efficiently detected and removed from the video frames.

Fig. 5 shows the configuration of the text frame detection and verification unit 302 shown in Fig. 3. The text frame detection and verification unit 302 includes a fast and simple image binarization unit 321, a text line region determination unit 322, a rebinarization unit 323, text line confirmation unit 324, and text frame verification unit 325. The fast and simple image binarization unit 321 generates the first binary image of a video frame. The text line region determination unit 322 determines the position of a text line region by using a horizontal projection and a vertical projection of the first binary image. The rebinarization unit 323 generates the second binary image of every text line region. The text line confirmation unit 324 determines the validity of a text line region by using the difference between the first binary image and the second binary image and a fill

rate of the number of foreground pixels in the text line region to the total number of pixels in the text line region. The text frame verification unit 325 confirms whether a set of continuous video frames are non-text frames that do not contain a text area by using the number of valid text line regions in the set of continuous video frames. According to the text frame detection and verification unit 302, non-text frames are fast detected and removed from the video frames.

Fig. 6 shows the configuration of the image shifting detection unit 303 shown in Fig. 3. The image shifting detection unit 303 includes a fast and simple image binarization unit 331, a text line vertical position determination unit 332, and a vertical shifting detection unit 333, a horizontal shifting detection unit 334. The fast and simple image binarization unit 331 generates binary images of two video frames. The text line vertical position determination unit 332 determines the vertical position of every text line region by using horizontal projections of the binary images. The vertical shifting detection unit 333 determines a vertical offset of image shifting between the two video frames and the similarity of the two video frames in the vertical direction by using the correlation between the horizontal projections. The horizontal shifting detection unit 334 determines a horizontal offset of the image shifting and the similarity of the two video frames in the horizontal direction by using the correlation between vertical projections of every text line in the binary images. According to the

image shifting detection unit 303, redundant frames caused by image shifting are fast detected and removed from the video frames.

Figs. 7 and 8 show two frames that have same text content. Fig. 9 shows the processing result of the frame similarity measurement unit 301 for these two frames. The white boxes in the fig. 9 mark out all valid image blocks which are blocks included by the valid block pairs and have the ability to show the change of the content. Boxes with solid line stand for similar image blocks and boxes with dashed line stand for dissimilar image blocks. Since the ratio of the number of similar image blocks to the number of valid blocks is larger than a predefined threshold, these two images are considered as similar and the second frame is removed.

Fig. 10 shows the flowchart of the operation of the frame similarity measurement unit 301 shown in Fig. 4. The comparison starts at 0 th frame of 0 th second (S501), the current i th frame is compared with the j th frame, which has a frame interval of STEP frames (S502). If the i th frame is similar with the j th frame in comparing the two frames (S503), then the current frame jumps to j th frame (S510) and the processes in S502 and S503 are repeated for comparison.

If the two frame are different, comparison restarts from one frame after the current frame, which is the k th frame (S504 and S505). It is checked whether k is less than j (S506). If the k th frame is before the j th frame, and if the i th frame is similar with the k th frame

(S511), then the current frame is assigned as the k th frame (S512), and the processes in S502 and S503 are repeated for comparison.

If the i th frame is different with the k th frame, then k increases by 1 (S505), and it is checked whether k is less than j. If k is not less than j, that means the j th frame is different with the previous frames, the j th frame is marked as a new candidate text change frame (S507). A new search begins from the j th frame (S508). If the sum of the index i of the current search frame and STEP is larger than the total number of input video frames nFrame (S509), then the search is over and the found candidate text change frames are sent to the following units 302 and 303 for further processing. Otherwise, the search is continued.

The purpose of the frame interval STEP is to reduce the total time for the search operation. If STEP is too big and the content of video changes rapidly, the performance will decrease. If the STEP is too small, the total search time will also be not very short. This frame interval is chosen as STEP = 4 frames, for example.

Fig. 11 shows the flowchart of the operation of the determination of the similarity of two frames in S503 shown in Fig. 10. The flowchart of the process in S511 is obtained by replacing j with k in Fig. 11.

At start, the image block count n, the valid block count nValid, and the similar block count nSimilar are all set to zero (S513). Then the i th frame and the j th frame are divided into non-overlapped small image blocks with

size of NxN, and the number of the image blocks is recorded as nBlock (S514). Here N = 16, for example. The two image blocks in the same position in the two frames are defined as an image block pair. For every image block pair, the image block validation unit 311 is used to check whether the image block pair is a valid block pair (S515). The detection of the change between two frames is achieved by detecting change in every image block pair. The background parts of a slide usually do not change, even if the content has been changed. So image block pairs in these parts should not considered as valid block pairs.

If the block pair is invalid, then the next block pair is checked (S519 and S520). If the block pair is a valid block pair, the valid block count nValid increases by 1 (S516), and the image block similarity measurement unit 312 is used to measure the similarity of the two image blocks (S517). If the blocks are similar, the similar block count nSimilar increases by 1 (S518). When all the block pairs are compared (S519 and S520), the frame similarity judgment unit 313 is used to determine whether the two frames are similar (S521). The two frames are considered as similar if the following condition is met (S522):

nSimilar > nValid * simrate,

here simrate = 0.85, for example. The two frames are considered as dissimilar if the above condition is not met (S523).

Fig. 12 shows the flowchart of the operation of the image block validation unit 311 in S515 shown in Fig. 11. First, the mean and the variance of the n th image block pair are calculated (S524). The means and the variances of the gray level of the image block in the i th frame are denoted by M(i) and V(i), respectively. The mean and the variance of the gray level of the image block in the j th frame are denoted by M(j) and V(j), respectively. If two variances V(i) and V(j) of the block pair are all smaller than a predefined threshold Tv (S525), and the absolute difference of the two means M(i) and M(j) is also smaller than a predefined threshold Tm (S526), then the image block pair is an invalid block pair (S527). Otherwise, the image block pair is a valid block pair (S528).

Fig. 13 shows the flowchart of the operation of the image block similarity measurement unit 312 in S517 shown in Fig. 11. The means M(i) and M(j) of the n th image block pair is calculated first (S529). If the absolute difference of the two means M(i) and M(j) is larger than a predefined threshold Tm1 (S530), then the two image blocks are considered as dissimilar image blocks (S534). Otherwise, the correlation of the two image blocks C(i, j) is calculated (S531). If the correlation C(i, j) is larger than a predefined threshold Tc (S532), the two image blocks are similar (S533), and if the correlation is smaller than the threshold Tc, the two image blocks are dissimilar (S534).

Figs. 14 to 21 show some example results of the processes performed by the text frame detection and

verification unit 302 shown in Fig. 5. Fig. 14 shows the original video frame. Fig. 15 shows the first binary image resulted from fast and simple image binarization. Fig. 16 shows the result of horizontal binary projection. Fig. 17 shows the result of projection regularization. Fig. 18 shows the result of vertical binary projection in every candidate text line. Fig. 19 shows the result of text line region determination. Gray rectangles indicate candidate text line regions.

Fig. 20 shows the result of two pairs of binary images for two candidate text line regions marked in dashed line in Fig. 19. The first pair binary images contain text information. The difference between these two images is very small. So this text line region is regarded as a true text line region. The second pair of binary images has very big difference. Since the different part is larger than a predefined threshold, this region is considered as non-text-line region. Fig. 21 shows the detected text line regions.

Figs. 22 and 23 show the flowchart of the operation of the text frame detection and verification unit 302 shown in Fig. 3. First, continuous candidate frames section detection is performed to classify the candidate text frames outputted by the frame similarity measurement unit 301 into a plurality of sections, each section contains a serial of continuous candidate frames (S701). The number of the sections is denoted by nSection. Started from the first section (S702), if the number of the continuous candidate frames M(i) of the i th section is

larger than a predefined threshold Tncf (S703), the fast and simple image binarization unit 321 is used to get the every binary image of all video frames (S704). Then the text line region determination unit 322 using the horizontal and vertical projection of the binary image is used to determine the regions of the text lines (S705).

Next started from the first detected text line region (S706), the rebinarization unit 323 is used to make a second binary image of the text line region (S707). The rebinarization unit 323 uses Niblack's image binarization method on the whole region of every detected text line to get the binary image. The two binary images of the same text line region are compared by the text line confirmation unit 324 (S708). If the two binary images are similar, then the similar text line count for the i th section nTextLine(i) increases by 1 (S709). This procedure repeat for all text lines in these M(i) continuous candidate frames (S710 and S711).

Sometime a non-text frame will be detected as containing some text lines, but if a serial of candidate frames do not contain any text line, it is unlikely that the total number of the text lines detected in these frames will be very big. So the text frame verification unit 325 is used to confirm whether the serial of candidate text frames are non-text frames. The serial of the candidate text frames are considered as non-text frames if the following condition is met (S712):

$$nTextLine(i) \leq \alpha M(i),$$

and these false candidate text frames are removed (S713). Here, α is a positive real number that is determined by experiment. Usually it is set as $\alpha = 0.8$. The procedure repeats for all continuous candidate frames sections (S714 and S715).

Fig. 24 shows the flowchart of the operation of the fast and simple binarization unit 321 in S704 shown in Fig. 22. The frame image is first divided into non-overlapped image blocks with size of N×N, and the number of the image blocks is recorded as nBlock (S716). Here N = 16, for example. Started from the first image block (S717), every image block is binarized using Niblack's image binariation method (S718). The parameter k for Niblack's image binarization is set as k = -0.4. The procedure repeats for all image blocks (S719 and S720).

Fig. 25 shows the flowchart of Niblack's image binarization method in S718 shown in Fig. 24. The input is a gray level image of size M×N. First, the mean Mean and the variance Var of the image is calculated (S721). If the variance Var is less than a predefined threshold Tv (S722), then all pixels in the binary image are set to 0. If Var > Tv, a binary threshold T is calculated by the following equation:

$$T = Mean + k * Var.$$

For every image pixel i, if the gray level gray(i) for of the pixel is larger than T (S726), the pixel in the

binary image bin(i) is set to 0 (S727), otherwise, the pixel is set to 1 (S728). The procedure repeats for all pixels in the binary image (S729 and S730).

Fig. 26 shows the flowchart of the operation of the text line region determination unit 322 in S705 shown in Fig. 22. The input of this unit is the binary image of the video frame from S704. The horizontal image projection Prjh is first calculated (S731). The projection is then smoothed (S732) and regularized (S733). The result of the regularization of Prjh is Prjhr, which has only two values: 0 or 1. 1 means that the position has a large projection value, 0 means that the position has a small projection value. The start and end points of each 1's region in the Prjhr are recorded as sy(i) and ey(i), respectively (S734). For each 1's region in Prjhr, the vertical image projection Prjv(i) is calculated (S735). Prjv(i) is smoothed (S736) and regularized as Prjvr(i) (S737). Two 1's regions in Prjvr(i) are connected into one region if the distance between the two 1's regions is less than 2 * region height, and the start and end points of the connected region are recorded as sx(i) and ex(i), respectively (S738). The output sx(i), ex(i), sy(i) and ey(i) determine the i th region of the text line (S739).

Fig. 27 shows the flowchart of horizontal image projection in S731 shown in Fig. 26. Started from the first horizontal line (S740), the projection for the i th horizontal line is calculated by the following equation (S741):

$$prj(i) = \sum_{j=0}^{w-1} I(i,j),$$

where I(i, j) is the pixel value in the i th row and j th column and w is the width of the image. The calculation repeats for all horizontal lines in the image with h as the height of the image (S742 and S743).

Fig. 28 shows the flowchart of projection smoothing in S732 shown in Fig. 26. Started from the radii of the smoothing window, δ (S744), the value for the i th point in the smoothed projection prjs(i) is calculated by the following equation (S745):

$$prjs(i) = \frac{1}{2\delta + 1} \sum_{j=i-\delta}^{i+\delta} prj(j),$$

where the length of the smoothing window is 2 * δ + 1. The calculation repeats for all points in the smocthed projection with L as the range for smoothing (S746 and S747).

Fig. 29 shows the flowchart of projection regularization in S733 shown in Fig. 26. At first, all local maxima in the projection are detected (S748). The value for every pixel in the regularized projection Prjr is set to 0 (S749). Started from the first local maximum max(i) (S750), two nearby local minima min1(i) and min2(i) are detected (S751).

Fig. 30 shows an exemplary drawing of the max(i), min1(i) and min2(i) positions in a projection curve. There are three local maxima. P2, P4 and P6 are max(1), max(2) and max(3), respectively. P1 is the upper minimum min1(1) for max(1), P3 is the bottom minimum min2(1) for max(1).

P3 is also the upper minimum min1(2) for max(2). Similarly, P5 is the bottom minimum min2(2) for max(2), and also is the upper minimum min1(3) for max(3). P7 is the bottom minimum min2(3) for max(3).

If min1(i) < max(i)/2 and min2(i) < max(i)/2 (S752), then the values in the regularized projection Prjr between the positions of min1(i) and min2(i) are set to 1 (S753). The procedure repeats for every local maximum (S754 and S755).

Fig. 31 shows the flowchart of the operation of the text line confirmation unit 324 in S708 shown in Fig. 22. The input of this unit is two binary images I1 and I2 with size wxh of the same text line region. First the counters count1, count2 and count are set to 0 (S756). count means the number of pixels where the value of two corresponding pixels in I1 and I2 are all 1. count1 means the number of pixels where the value of the corresponding pixel in I1 is 1 and that in I2 is 0. count2 means the number of pixels where the value of the corresponding pixel in I2 is 1 and that in I1 is 0.

Started from the first position in the two images, if corresponding two pixels I1(i) and I2(i) are both 1, then count increases by 1 (S757 and S758). Otherwise, if I1(i) is 1, then count1 increases by 1 (S759 and S760). Otherwise, if I2(i) is 1, then count2 increases by 1 (S761 and S762). After all pixels are checked (S763 and S764), it is checked whether the following conditions are met (S765 and S766):

count + count1 < w * h/2,

count + count2 < w * h/2,

count1 < count * 0.2,

count2 < count * 0.2,

fillrate < 0.5.

The 'fillrate' of a text line region is defined as the rate of the number of foreground pixels to the number of total pixels in the region. If the above conditions are met, then two binary images are considered as similar in this text line region and the text line region is considered as a valid text line (S768). If one of these conditions is not met, the text line region is considered as an invalid text line (S767).

Figs. 32 and 33 show the flowchart of the operation of the image shifting detection unit 303 shown in Fig. 6. For two continuous frames, frame i and frame j, first the fast and simple image binarization unit 331 is used to get the binary image of the two frames (S801). Then the text line vertical position determination unit 332 is used to perform the horizontal image projection as in S731 shown in Fig. 26 for obtaining the horizontal projections Prjyi and Prjyj for frame i and frame j, respectively (S802). The vertical shifting detection unit 333 is then used to calculate the correlation function Cy(t) of the two projections (S803).

Here, a correlation function C(t) of two projections Prj1(x) and Prj2(x) is defined as:

$$C(t)=\frac{1}{L*V1*V2}\sum(\mathrm{Prj1}(x)-\mathrm{M1})*(\mathrm{Prj2}(x+t)-\mathrm{M2})$$

Where L is the length of the projection, and M1 and M2 are the means of the projections Prj1 and Prj2, respectively. V1 and V2 are the variances of Prj1 and Prj2, respectively.

If the maximum of Cy(t) is less than 90% (S804), then the two images are not shifting images. Otherwise, the position of the maximum value of Cy(t) is recorded as the vertical offset offy (S805), and the projection regularization as in S733 is performed to get the regularized projection Prjyir of projection Prjyi (S806). If frame j is a shifting version of frame i, the vertical shifting offset of frame j is represented by offy. Every 1's region in Prjyir is considered as a candidate text line region, which is indicated by the start and end points syi and eyi (S807). The number of the candidate text line regions is recorded as nCanTL.

Started from the first candidate text line region, the matching count nMatch is set to 0 (S808). The c th corresponding shifting candidate text line region in frame j is assumed to be represented by syj(c) = syi(c) + offy and eyj(c) = eyi(c) + offy (S809). For two corresponding candidate text line regions, the vertical projections are calculated (S810). Then the horizontal shifting detection unit 334 is used to calculate the correlation function Cx(t) for the two vertical projections is calculated, and the position of the maximum value of Cx(t) is recorded as the horizontal offset offx, for these two projections (S811). If the maximum of Cx(t) is larger than 90% (S812),

the two candidate text line regions are considered as matched shifting text line regions and the matching count nMatch increases by 1 (S813). After every candidate text line pair are checked (S814 and S815), if the number of the matched shifting text line regions is larger than 70% of the number of candidate text line regions (S816), frame j is regarded as a shifting version of frame i (S817). Otherwise, frame j is not a shifting frame of frame i (S818).

Fig. 34 shows the configuration of the text extraction apparatus 105 shown in Fig. 1. The text extraction apparatus comprises an edge image generation unit 901 for extracting the edge information of the video frame, a stroke image generation unit 902 using the edge image for generating the stroke image of the candidate character strokes, a stroke filtering unit 903 for removing false character strokes, a text line region formation unit 904 for connecting nearby strokes into a text line region, a text line verification unit 905 for delete false character stroke in the text line region, and a text line binarization unit 906 for obtaining the final binary image of the text line region. The output of the text extraction apparatus is a list of binary images for all text line regions in the frame. According to the text extraction apparatus 105, the text line region can be accurately binarized since the false strokes are detected and removed as much as possible.

Fig. 35 shows the configuration of the edge image generation unit 901 shown in Fig. 34. The edge image

generation unit 901 includes an edge strength calculation unit 911, a first edge image generation unit 912, and a second edge image generation unit 913. The edge strength calculation unit 911 calculates edge strength for every pixel in a video frame by using a Sobel edge detector. The first edge image generation unit 912 generates the first edge image by comparing the edge strength of every pixel with a predefined edge threshold and sets a value of a corresponding pixel in the first edge image to one binary value if the edge strength is greater than the threshold and the other binary value if the edge strength is less than the threshold. For example, logic "1" is used as the one binary value, which may indicate a white pixel, and logic "0" is used as the other binary value, which may indicate a black pixel. The second edge image generation unit 913 generates the second edge image by comparing the edge strength of every pixel in a window centered at the position of every pixel of the one binary value in the first edge image with mean edge strength of the pixels in the window, and sets a value of a corresponding pixel in the second edge image to the one binary value if the edge strength of the pixel is greater than the mean edge strength and the other binary value if the edge strength of the pixel is less than the mean edge strength. A small window of size of 3×3, for example, is used for the second edge image generation.

Fig. 36 shows the configuration of the stroke image generation unit 902 shown in Fig. 34. The stroke image generation unit 902 includes a local image binarization

unit 921. The local image binarization unit 921 binarizes a gray scale image of the video frame in the Niblack's binarization method to obtain a binary image of candidate character strokes by using a window centered at the position of every pixel of the one binary value in the second edge image. A window of size of 11×11, for example, is used for the local image binarization.

Fig. 37 shows the configuration of the stroke filtering unit 903 shown in Fig. 34. The stroke filtering unit 903 includes a stroke edge coverage validation unit 931 and a long straight line detection unit 932. The stroke edge coverage validation unit 931 checks an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels of the one binary value in the second edge image, determines that the stroke is a valid stroke if the overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold, and removes the invalid stroke as a false stroke. The long straight line detection unit 932 removes a very large stroke as a false stroke by using a width and a height of the stroke. According to the stroke filtering unit 903, false strokes unnecessary for a text line region are detected and removed from the binary image of the candidate character strokes.

Fig. 38 shows the configuration of the text line region formation unit 904 shown in Fig. 34. The text line region formation unit 904 includes a stroke connection checking unit 941. The stroke connection checking unit 941

checks whether two adjacent strokes are connectable by using an overlap ratio of heights of the two strokes and a distance between the two strokes. The text line region formation unit 904 combines strokes into a text line region by using the result of the checking.

Fig. 39 shows the configuration of the text line verification unit 905 shown in Fig. 34. The text line verification unit 905 includes a vertical false stroke detection unit 951, a horizontal false stroke detection unit 952, and a text line reformation unit 953. The vertical false stroke detection unit 951 checks every stroke with a height higher than the mean height of strokes in the text line region, and marks the stroke as a false stroke if the stroke connects two horizontal text line regions into one big text line region. The horizontal false stroke detection unit 952 checks every stroke with a width larger than a threshold determined by the mean width of the strokes in the text line region, and marks the stroke as a false stroke if the number of strokes in a region that contains the stroke is less than a predefined threshold. The text line reformation unit 953 reconnects strokes except for a false stroke in the text line region if the false stroke is detected in the text line region. According to the text line verification unit 905, false strokes are further detected and removed from the text line region.

Fig. 40 shows the configuration of the text line binarization unit 906 shown in Fig. 34. The text line binarization unit 906 includes an automatic size

calculation unit 961 and a block image binarization unit 962. The automatic size calculation unit 961 determines a size of a window for binarization. The block image binarization unit 962 binarizes a gray scale image of the video frame in the Niblack's binarization method to obtain a binary image of a text line region by using the window centered at the position of every pixel of the one binary value in the second edge image. According to such text line binarization after removing the false strokes, the text line region can be accurately binarized.

Figs. 41 to 46 show some results of the text extraction apparatus. Fig. 41 shows the original video frame. Fig. 42 shows the result for edge image generation, which is the final edge image (second edge image). Fig. 43 shows the result of stroke generation. Fig. 44 shows the result of stroke filtering. Fig. 45 shows the result of text line formation. Fig. 46 shows the result of the refined final binarized text line regions.

Figs. 47 and 48 show the flowchart of the operation of the edge image generation unit 901 shown in Fig. 35. First all the values of the pixels EdgeImg1(i) in the first edge image EdgeImg1 of size W×H are set to 0 (S1101). Started from the first pixel (S1102), the edge strength calculation unit 911 is then used to calculate edge strength E(i) of the i th pixel using Sobel edge detector (S1103). Next the first edge image generation unit 912 is used to determine the value of EdgeImg1(i). If the edge strength is larger than a predefined threshold Tedge (S1104), then the value of this pixel in the first edge

image is set to 1, EdgeImg1(i) = 1 (S1105). This procedure continues until all the pixels are checked (S1106 and S1107).

After the first edge image is obtained, all the values EdgeImg2(i) for the second edge image EdgeImg2 of size W×H are initialized to 0 (S1108). Scanned from the first pixel (S1109), if the value of the pixel in the first edge image is 1 (S1110), then the mean edge strength of the neighborhood pixels is obtained according to the arrangement of the neighborhood 1116 of pixel i shown in Fig. 49 (S1111). The second edge image generation unit 913 is then used to determine the values for these neighborhood pixels in the second edge image by comparing the edge strength of a pixel with the mean edge strength (S1112). If the edge strength is larger than the mean edge strength, the pixel value in the second edge image is set to 1, otherwise the value is set to 0. After all pixels in the first edge image are checked (S1113 and S1114), the second edge image is outputted as the final edge image EdgeImg (S1115).

Fig. 50 shows the flowchart of the operation of the edge strength calculation unit 911 in S1103 shown in Fig. 47. For the i th pixel, the horizontal and the vertical edge strengths Ex(i) and Ey(i) are first obtained in the neighborhood area 1116 shown in Fig. 49 by the following equations (S1117 and S1118):

$$Ex(i) = I(d) + 2*I(e) + I(f) - I(b) - 2*I(a) - I(h),$$

$$Ey(i) = I(b) + 2*I(c) + I(d) - I(h) - 2*I(g) - I(f),$$

where I(x) represents the gray level of the x th pixel (x = a, b, c, d, e, f, g, h). The total edge strength E(i) is calculated by the following equation (S1119):

$$E(i) = \sqrt{Ex(i) * Ex(i) + Ey(i) * Ey(i)} .$$

The mean edge strength of pixel i in S1111 shown in Fig. 48 is calculated by the following equation:

Medge(i) = (E(a) + E(b) + E(c) + E(d) + E(e) + E(f) + E(g) + E(h) + E(i)) / 9.

Fig. 51 shows the flowchart of the operation of the stroke image generation unit 902 shown in Fig. 36. The stroke image of size W×H is first initialized to 0 (S1201). Then the local image binarization unit 921 is used to determine the values of the pixels of the stroke image. Started from the first pixel (S1202), if the value of the i th pixel EdgeImg(i) in the edge image EdgeImg is 1 (S1203), a 11×11 window is set at the gray level frame image centered at the pixel's position and the values of the pixels of the stroke image in the window are determined by Niblack's binarization method shown in Fig. 25 (S1204). After all pixels are checked in the edge image (S1205 and S1206), the stroke image is generated.

Fig. 52 shows the flowchart of the operation of the stroke filtering unit 903 shown in Fig. 37. First the long straight line detection unit 932 is used to delete very

large strokes. Started from the first stroke (S1301), if the width or height of the stroke exceeds a predefined threshold MAXSTROKESIZE (S1302), this stroke is deleted (S1304). Otherwise, the stroke edge coverage validation unit 931 is used to check the validity of the stroke (S1303). A valid stroke means a candidate character stroke and an invalid stroke is not a true character stroke. If the stroke is invalid, it is deleted (S1304). The checking is repeated for all strokes found in the stroke image with nStroke as the number of the strokes (S1305 and S1306).

Fig. 53 shows the flowchart of the operation of the stroke edge coverage validation unit 931 in S1303 shown in Fig. 52. First the contour C of the stroke is obtained (S1307). From the first contour point (S1308), the pixel values of EdgeImg in the neighborhood area of the current contour point are checked (S1309). As shown in Fig. 49, point a to point h are considered as the neighborhood points of point i. If there is a neighbor edge pixel which has a value of 1, then this contour point is regarded as a valid edge contour point and the count of valid edge contour points nEdge increases by 1 (S1310). After all contour points are checked with nContour as the number of the contour points (S1311 and S1312), if the number of the valid edge contour points is larger than 0.8*nContour (S1313), the stroke is considered as a valid stroke, that is, a candidate character stroke (S1314). Otherwise, the stroke is an invalid stroke (S1315). An invalid stroke is deleted from the stroke list. The rate of nEdge to nContour in S1313 represents the overlap rate.

Fig. 54 shows the flowchart of the operation of the text line region formation unit 904 shown in Fig. 38. First the region of every stroke is set as an individual text line region and the number of text line nTL is set to nStroke (S1401). Started from the first stroke (S1402), stroke j next to stroke i is selected (S1403) and it is checked whether stroke i and stroke j belong to one text line region (S1404). If not, the stroke connection checking unit 941 is used to check whether these two strokes are connectable (S1405). If so, all the strokes in these two text lines, a text line to which stroke i belongs and a text line to which stroke j belongs, are combined into one big text line (S1406) and the number of text line decreases by 1 (S1407).

Here, a text line is a group of connectable strokes and every stroke has an attribute of a text line. If stroke i belongs to the m th text line, stroke j belongs to the n th text line, and stroke i is connectable with stroke j, then the attributes of all strokes in the m th and the n th text lines are set to m. After every pair of the strokes are checked (S1408, S1409, S1410 and S1411), nTL is the number of the text lines in the frame.

Fig. 55 shows the flowchart of the operation of the stroke connection checking unit 941 in S1405 shown in Fig. 54. First, the heights of the two strokes h1 and h2 are obtained and the higher height is marked as maxh and the lower height is marked as minh (S1412). If the horizontal distance between the centers of stroke i and stroke j is larger than 1.5*maxh (S1413), then these two strokes are

not connectable (S1417). Otherwise, the number of the horizontal lines that has intersection with both stroke i and stroke j is recorded as nOverlap (S1414). If nOverlap is larger than 0.5*minh (S1415), then these two strokes are connectable (S1416). Otherwise, these two strokes are not connectable (S1417). The ratio of nOverlap to minh in S1415 represents the overlap ratio.

Fig. 56 shows the flowchart of the operation of the text line verification unit 905 shown in Fig. 39. First, the modification flag modflag is set to false (S1501). Started from the first text line region (S1502), if the height of the i th text line region Height(i) is less than a predefined threshold MINTLHEIGHT (S1503), this text line region is deleted (S1504). Otherwise, a vertical false stroke detection unit 951 and a horizontal false stroke detection unit 952 are used to detect a false stroke (S1505 and S1506). If a false stroke is detected, then the stroke is deleted (S1507), the remaining strokes are reconnected (S1508) using the text line reformation unit 953, and the modification flag is set to true (S1509). The text line reformation unit 953 reconnects the remaining strokes in the same manner as the text line region formation unit 904. After all the text line regions are checked (S1510 and S1511), if the modification flag is true (S1512), then the whole process is repeated again until no false stroke is detected.

Fig. 57 shows the flowchart of the operation of the vertical false stroke detection unit 951 in S1505 shown in Fig. 56. The mean height of the strokes in the text line

region is first calculated (S1513). Started from the first stroke (S1514), if the height of stroke i is larger than the mean height (S1515), then multiple text line detection is performed to check the strokes in an area to the left of stroke i (S1516). The area to the left of stroke i is a region inside a text line region, and the left, up, and bottom boundaries of this area are the left, up and bottom boundaries, respectively, of the text line region. The right boundary of this area is the left boundary of stroke i. If there are two or more non-overlapped horizontal text line regions in the area to the left of stroke i, stroke i is a vertical false stroke (S1520).

Otherwise, multiple text line detection is then performed to check the strokes in an area to the right of stroke i (S1517). The area to the right of stroke i has a similar definition to that of the area to the left of stroke i. If there are two or more non-overlapped horizontal text line regions in the area to the right of stroke i, stroke i is a vertical false stroke (S1520). The procedure repeats until every stroke in the text line region is checked (S1518 and S1519).

Fig. 58 shows the flowchart of multiple text line detection in S1516 and S1517 shown in Fig. 57. First, the strokes are connected in the same manner as the text region formation unit 904 (S1521). If the number of the text line regions nTextLine is more than 1 (S1522), then it is checked whether the following three conditions are met.

1. There are two non-overlapped text line regions (S1523).

2. One text line region is above the other text line region (S1524).

3. Number of the strokes in each text line region is larger than 3 (S1525).

If all the three conditions are met, then multiple text lines are detected (S1526).

Fig. 59 shows the flowchart of the operation of the horizontal false stroke detection unit 952 in S1506 shown in Fig. 56. First, the mean width of all the strokes in the text line region is calculated (S1527). Started from the first stroke (S1528), if the width of stroke i is larger than 2.5 times the mean stroke width (S1529), then a detection region R is set (S1530). The left boundary R.left and the right boundary R.right of R are determined by the left boundary Stroke(i).Left and the right boundary Stroke(i).Right, respectively, of stroke i. The top boundary R.top and the bottom boundary R.bottom of R are determined by the top boundary textline.top and the bottom boundary textline.bottom, respectively, of the text line region. The number of strokes in detection region R is calculated (S1531), if the number is less than or equal to 3 (S1532), then stroke i is marked as a horizontal false stroke (S1533). The procedure repeats until every stroke in the text line region is checked (S1534 and S1535).

Figs. 60 and 61 show examples of a false stroke. Stroke 1541 shown in Fig. 60 is a vertical false stroke and stroke 1542 shown in Fig. 61 is a horizontal false stroke.

Fig. 62 shows the flowchart of the operation of the

text line binarization unit 906 shown in Fig. 40. First, the automatic size calculation unit 961 is used to determine the size of the window wh for binarization based on the height of the text line region Height (S1601), which must satisfy the following three conditions:

wh = Height / 3,
wh = wh + 1 if wh is an even number,
wh = 5 if wh < 5.

After that, the block image binarization unit 962 is used to rebinarize the text line region (S1602). The block image binarization unit 962 sets the window size of Niblack's binarization method to wh and rebinarizes the text line region in the same manner as the stroke image generation unit 902.

The video text processing apparatus or each of the text change frame detection apparatus 104 and the text extraction apparatus 105 shown in Fig. 1 is configured, for example, using an information processing apparatus (computer) as shown in Fig. 63. The information processing apparatus shown in Fig. 63 comprises a CPU (central processing device) 1701, a memory 1702, an input device 1703, an output device 1704, an external storage device 1705, a medium drive device 1706, a network connection device 1707, and a video input device 1708. They are interconnected through a bus 1709.

The memory 1702 includes, for example, ROM (read only memory), RAM (random access memory), etc. and stores

programs and data for use in the processes. The CPU 1701 performs a necessary process by executing the program using the memory 1702. In this case, the units 301 to 303 shown in Fig. 3 and the units 901 to 906 shown in Fig. 34 correspond to the programs stored in the memory 1702.

The input device 1703 is, for example, a keyboard, a pointing device, a touch panel, etc., and used to input an instruction and information from a user. The output device 1704 is, for example, a display, a printer, a speaker, etc., and used to output an inquiry to the user and a process result.

The external storage device 1705 is, for example, a magnetic disk device, an optical disk device, a magneto-optical disk device, a tape device, etc. The information processing apparatus stores the programs and data in the external storage device 1705, and loads them to the memory 1702 to use them as necessary. The external storage device 1705 is also used as a database storing the existing video data 101 shown in Fig. 1.

The medium drive device 1706 drives a portable storage medium 1710, and accesses the stored contents. The portable storage medium 1710 is an arbitrary computer-readable storage medium such as a memory card, a flexible disk, CD-ROM (compact disk read only memory), an optical disk, a magneto-optical disk, etc. The user stores the programs and data in the portable storage medium 1710, and loads them to the memory 1702 to use them as necessary.

The network connection device 1707 is connected to an arbitrary communications network such as a LAN (local

area network), Internet, etc., and converts data during the communications. The information processing apparatus receives the programs and data through the network connection device 1707, loads them to the memory 1702 to use them as necessary.

The video input device 1708 is, for example, the TV video camera 102 shown in Fig. 1 and is used to input the living video stream.

Fig. 64 shows computer-readable storage media capable of providing a program and data for the information processing apparatus shown in Fig. 63. The program and data stored in the portable storage medium 1710 and a database 1803 of a server 1801 are loaded to the memory 1702 of an information processing apparatus 1802. The server 1801 generates a propagation signal for propagating the program and data, and transmits it to the information processing apparatus 1802 through an arbitrary transmission medium in a network. The CPU 1701 executes the program using the data to perform a necessary process.

(addition 1) A text change frame detection apparatus that selects a plurality of video frames including text contents from given video frames, characterized in that said apparatus comprises:

first frame removing means for removing redundant video frames from the given video frames;

second frame removing means for removing video frames that do not contain a text area from the given video frames;

third frame removing means for detecting and

removing redundant video frames caused by image shifting from the given video frames; and

output means for outputting remaining video frames as candidate text change frames.

(addition 2) The text change frame detection apparatus according to addition 1, characterized in that the first frame removing means includes:

image block validation means for determining whether two image blocks in the same position in two video frames of the given video frames are a valid block pair that has an ability to show a change of image contents;

image block similarity measurement means for calculating a similarity of two image blocks of the valid block pair and determining whether the two image blocks are similar; and

frame similarity judgment means for determining whether the two video frames are similar by using a ratio of a number of similar image blocks to a total number of valid block pairs,

and the first frame removing means removes a similar video frame as a redundant video frame.

(addition 3) The text change frame detection apparatus according to addition 1, characterized in that the second frame removing means includes:

fast and simple image binarization means for generating a first binary image of a video frame of the given video frames;

text line region determination means for determining a position of a text line region by using a horizontal

projection and a vertical projection of the first binary image;

rebinarization means for generating a second binary image of every text line region;

text line confirmation means for determining validity of a text line region by using a difference between the first binary image and the second binary image and a fill rate of a number of foreground pixels in the text line region to a total number of pixels in the text line region; and

text frame verification means for confirming whether a set of continuous video frames are non-text frames that do not contain a text area by using a number of valid text line regions in the set of continuous video frames.

(addition 4) The text change frame detection apparatus according to addition 1, characterized in that the third frame removing means includes:

fast and simple image binarization means for generating binary images of two video frames of the given video frames;

text line vertical position determination means for determining a vertical position of every text line region by using horizontal projections of the binary images of the two video frames;

vertical shifting detection means for determining a vertical offset of image shifting between the two video frames and a similarity of the two video frames in a vertical direction by using correlation between the horizontal projections; and

horizontal shifting detection means for determining a horizontal offset of the image shifting and a similarity of the two video frames in a horizontal direction by using correlation between vertical projections of every text line in the binary images of the two video frames,
and the third frame removing means removes a similar video frame as a redundant video frame caused by the image shifting.

(addition 5)　A text change frame detection apparatus that selects a plurality of video frames including text contents from given video frames, characterized in that said apparatus comprises:

image block validation means for determining whether two image blocks in the same position in two video frames of given video frames are a valid block pair that has an ability to show a change of image contents;

image block similarity measurement means for calculating a similarity of two image blocks of the valid block pair and determining whether the two image blocks are similar;

frame similarity judgment means for determining whether the two video frames are similar by using a ratio of a number of similar image blocks to a total number of valid block pairs; and

output means for outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

(addition 6)　A text change frame detection apparatus that selects a plurality of video frames including text

contents from given video frames, characterized in that said apparatus comprises:

fast and simple image binarization means for generating a first binary image of a video frame of the given video frames;

text line region determination means for determining a position of a text line region by using a horizontal projection and a vertical projection of the first binary image;

rebinarization means for generating a second binary image of every text line region;

text line confirmation means for determining validity of a text line region by using a difference between the first binary image and the second binary image and a fill rate of a number of foreground pixels in the text line region to a total number of pixels in the text line region;

text frame verification means for confirming whether a set of continuous video frames are non-text frames that do not contain a text area by using a number of valid text line regions in the set of continuous video frames; and

output means for outputting remaining video frames after the non-text frames are removed, as candidate text change frames.

(addition 7) A text change frame detection apparatus that selects a plurality of video frames including text contents from given video frames, characterized in that said apparatus comprises:

fast and simple image binarization means for

generating binary images of two video frames of the given video frames;

text line vertical position determination means for determining a vertical position of every text line region by using horizontal projections of the binary images of the two video frames;

vertical shifting detection means for determining a vertical offset of image shifting between the two video frames and a similarity of the two video frames in a vertical direction by using correlation between the horizontal projections;

horizontal shifting detection means for determining a horizontal offset of the image shifting and a similarity of the two video frames in a horizontal direction by using correlation between vertical projections of every text line in the binary images of the two video frames; and

output means for outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

(addition 8) A text extraction apparatus that extracts at least one text line region from a given image, characterized in that said apparatus comprises:

edge image generation means for generating edge information of the given image;

stroke image generation means for generating a binary image of candidate character strokes in the given image by using the edge information;

stroke filtering means for removing a false stroke from the binary image by using the edge information;

text line region formation means for combining a plurality of strokes into a text line region;

text line verification means for removing a false character stroke from the text line region and reforming the text line region;

text line binarization means for binarizing the text line region by using a height of the text line region; and

output means for outputting a binary image of the text line region.

(addition 9) The text extraction apparatus according to addition 8, characterized in that the edge image generation means includes:

edge strength calculation means for calculating edge strength for every pixel in the given image by using a Sobel edge detector;

first edge image generation means for generating a first edge image by comparing the edge strength of every pixel with a predefined edge threshold and setting a value of a corresponding pixel in the first edge image to one binary value if the edge strength is greater than the threshold and the other binary value if the edge strength is less than the threshold; and

second edge image generation means for generating a second edge image by comparing the edge strength of every pixel in a window centered at a position of every pixel of the one binary value in the first edge image with mean edge strength of the pixels in the window and setting a value of a corresponding pixel in the second edge image to the one binary value if the edge strength of the pixel is

greater than the mean edge strength and the other binary value if the edge strength of the pixel is less than the mean edge strength.

(addition 10)　The text extraction apparatus according to addition 9, characterized in that the stroke image generation means includes a local image binarization means for binarizing a gray scale image of the given image in a Niblack's binarization method to obtain the binary image of the candidate character strokes by using a window centered at a position of every pixel of the one binary value in the second edge image.

(addition 11)　The text extraction apparatus according to addition 9, characterized in that the stroke filtering means includes:

stroke edge coverage validation means for checking an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels of the one binary value in the second edge image, determining that the stroke is a valid stroke if the overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold, and removing the invalid stroke; and

long straight line detection means for removing a large stroke by using a width and a height of the stroke.

(addition 12)　The text extraction apparatus according to addition 9, characterized in that the text line binarization means includes:

automatic size calculation means for determining a size of a window for binarization; and

block image binarization means for binarizing a gray scale image of the given image in a Niblack's binarization method to obtain the binary image of the text line region by using the window centered at a position of every pixel of the one binary value in the second edge image.

(addition 13)　The text extraction apparatus according to addition 8, characterized in that the text line region formation means includes a stroke connection checking means for checking whether two adjacent strokes are connectable by using an overlap ratio of heights of the two strokes and a distance between the two strokes, and the text line region formation means combines the plurality of strokes into a text line region by using a result of checking.

(addition 14)　The text extraction apparatus according to addition 8, characterized in that the text line verification means includes:

vertical false stroke detection means for checking every stroke with a height higher than a mean height of strokes in the text line region, and marking the stroke as a false stroke if the stroke connects two horizontal text line regions into one big text line region;

horizontal false stroke detection means for checking every stroke with a width larger than a threshold determined by a mean width of the strokes in the text line region, and marking the stroke as a false stroke if a number of strokes in a region that contains the stroke is less than a predefined threshold; and

text line reformation means for reconnecting strokes

except for a false stroke in the text line region if the false stroke is detected in the text line region.

(addition 15)　A text extraction apparatus that extracts at least one text line region from a given image, characterized in that said apparatus comprises:

edge image generation means for generating an edge image of the given image;

stroke image generation means for generating a binary image of candidate character strokes in the given image by using the edge image;

stroke filtering means for checking an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels indicating an edge in the edge image, determining that the stroke is a valid stroke if the overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold, and removing the invalid stroke; and

output means for outputting information of remaining strokes in the binary image of the candidate character strokes.

(addition 16)　A program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

removing redundant video frames from the given video frames;

removing video frames that do not contain a text area from the given video frames;

detecting and removing redundant video frames caused by image shifting from the given video frames; and

outputting remaining video frames as candidate text change frames.

(addition 17) The program according to addition 16, characterized in that the program directs the computer to perform a process comprising:

determining whether two image blocks in the same position in two video frames of the given video frames are a valid block pair that has an ability to show a change of image contents;

calculating a similarity of two image blocks of the valid block pair and determining whether the two image blocks are similar;

determining whether the two video frames are similar by using a ratio of a number of similar image blocks to a total number of valid block pairs; and

removing a similar video frame as a redundant video frame.

(addition 18) The program according to addition 16, characterized in that the program directs the computer to perform a process comprising:

generating a first binary image of a video frame of the given video frames;

determining a position of a text line region by using a horizontal projection and a vertical projection of the first binary image;

generating a second binary image of every text line region;

determining validity of a text line region by using a difference between the first binary image and the second binary image and a fill rate of a number of foreground pixels in the text line region to a total number of pixels in the text line region; and

confirming whether a set of continuous video frames are non-text frames that do not contain a text area by using a number of valid text line regions in the set of continuous video frames.

(addition 19) The program according to addition 16, characterized in that the program directs the computer to perform a process comprising:

generating binary images of two video frames of the given video frames;

determining a vertical position of every text line region by using horizontal projections of the binary images of the two video frames;

determining a vertical offset of image shifting between the two video frames and a similarity of the two video frames in a vertical direction by using correlation between the horizontal projections; and

determining a horizontal offset of the image shifting and a similarity of the two video frames in a horizontal direction by using correlation between vertical projections of every text line in the binary images of the two video frames,

and the detecting and removing redundant video frames removes a similar video frame as a redundant video frame caused by the image shifting.

(addition 20)　A program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

determining whether two image blocks in the same position in two video frames of given video frames are a valid block pair that has an ability to show a change of image contents;

calculating a similarity of two image blocks of the valid block pair and determining whether the two image blocks are similar;

determining whether the two video frames are similar by using a ratio of a number of similar image blocks to a total number of valid block pairs; and

outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

(addition 21)　A program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

generating a first binary image of a video frame of the given video frames;

determining a position of a text line region by using a horizontal projection and a vertical projection of the first binary image;

generating a second binary image of every text line region;

determining validity of a text line region by using a difference between the first binary image and the second

binary image and a fill rate of a number of foreground pixels in the text line region to a total number of pixels in the text line region;

confirming whether a set of continuous video frames are non-text frames that do not contain a text area by using a number of valid text line regions in the set of continuous video frames; and

outputting remaining video frames after the non-text frames are removed, as candidate text change frames.

(addition 22) A program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

generating binary images of two video frames of the given video frames;

determining a vertical position of every text line region by using horizontal projections of the binary images of the two video frames;

determining a vertical offset of image shifting between the two video frames and a similarity of the two video frames in a vertical direction by using correlation between the horizontal projections;

determining a horizontal offset of the image shifting and a similarity of the two video frames in a horizontal direction by using correlation between vertical projections of every text line in the binary images of the two video frames; and

outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

(addition 23)　A program for a computer that extracts at least one text line region from a given image, characterized in that the program directs the computer to perform a process comprising:

generating edge information of the given image;

generating a binary image of candidate character strokes in the given image by using the edge information;

removing a false stroke from the binary image by using the edge information;

combining a plurality of strokes into a text line region;

removing a false character stroke from the text line region and reforming the text line region;

binarizing the text line region by using a height of the text line region; and

outputting a binary image of the text line region.

(addition 24)　The program according to addition 23, characterized in that the program directs the computer to perform a process comprising:

calculating edge strength for every pixel in the given image by using a Sobel edge detector;

generating a first edge image by comparing the edge strength of every pixel with a predefined edge threshold and setting a value of a corresponding pixel in the first edge image to one binary value if the edge strength is greater than the threshold and the other binary value if the edge strength is less than the threshold; and

generating a second edge image by comparing the edge strength of every pixel in a window centered at a position

of every pixel of the one binary value in the first edge image with mean edge strength of the pixels in the window and setting a value of a corresponding pixel in the second edge image to the one binary value if the edge strength of the pixel is greater than the mean edge strength and the other binary value if the edge strength of the pixel is less than the mean edge strength.

(addition 25) The program according to addition 24, characterized in that the program directs the computer to perform a process of binarizing a gray scale image of the given image in a Niblack's binarization method to obtain the binary image of the candidate character strokes by using a window centered at a position of every pixel of the one binary value in the second edge image.

(addition 26) The program according to addition 24, characterized in that the program directs the computer to perform a process comprising:

removing a large stroke by using a width and a height of the stroke;

checking an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels of the one binary value in the second edge image;

determining that the stroke is a valid stroke if the overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold; and

removing the invalid stroke.

(addition 27) The program according to addition 24, characterized in that the program directs the computer to

perform a process comprising:

determining a size of a window for binarization; and

binarizing a gray scale image of the given image in a Niblack's binarization method to obtain the binary image of the text line region by using the window centered at a position of every pixel of the one binary value in the second edge image.

(addition 28) The program according to addition 23, characterized in that the program directs the computer to perform a process comprising:

checking whether two adjacent strokes are connectable by using an overlap ratio of heights of the two strokes and a distance between the two strokes; and

combining the plurality of strokes into a text line region by using a result of checking.

(addition 29) The program according to addition 23, characterized in that the program directs the computer to perform a process comprising:

checking every stroke with a height higher than a mean height of strokes in the text line region;

marking the stroke as a false stroke if the stroke connects two horizontal text line regions into one big text line region;

checking every stroke with a width larger than a threshold determined by a mean width of the strokes in the text line region;

marking the stroke as a false stroke if a number of strokes in a region that contains the stroke is less than a predefined threshold; and

reconnecting strokes except for a false stroke in the text line region if the false stroke is detected in the text line region.

(addition 30)　A program for a computer that extracts at least one text line region from a given image, characterized in that the program directs the computer to perform a process comprising:

generating an edge image of the given image;

generating a binary image of candidate character strokes in the given image by using the edge image;

checking an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels indicating an edge in the edge image;

determining that the stroke is a valid stroke if the overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold;

removing the invalid stroke; and

outputting information of remaining strokes in the binary image of the candidate character strokes.

(addition 31)　A computer-readable storage medium storing a program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

removing redundant video frames from the given video frames;

removing video frames that do not contain a text area from the given video frames;

detecting and removing redundant video frames caused by image shifting from the given video frames; and

outputting remaining video frames as candidate text change frames.

(addition 32) A computer-readable storage medium storing a program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

determining whether two image blocks in the same position in two video frames of given video frames are a valid block pair that has an ability to show a change of image contents;

calculating a similarity of two image blocks of the valid block pair and determining whether the two image blocks are similar;

determining whether the two video frames are similar by using a ratio of a number of similar image blocks to a total number of valid block pairs; and

outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

(addition 33) A computer-readable storage medium storing a program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

generating a first binary image of a video frame of the given video frames;

determining a position of a text line region by

using a horizontal projection and a vertical projection of the first binary image;

generating a second binary image of every text line region;

determining validity of a text line region by using a difference between the first binary image and the second binary image and a fill rate of a number of foreground pixels in the text line region to a total number of pixels in the text line region;

confirming whether a set of continuous video frames are non-text frames that do not contain a text area by using a number of valid text line regions in the set of continuous video frames; and

outputting remaining video frames after the non-text frames are removed, as candidate text change frames.

(addition 34) A computer-readable storage medium storing a program for a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

generating binary images of two video frames of the given video frames;

determining a vertical position of every text line region by using horizontal projections of the binary images of the two video frames;

determining a vertical offset of image shifting between the two video frames and a similarity of the two video frames in a vertical direction by using correlation between the horizontal projections;

determining a horizontal offset of the image shifting and a similarity of the two video frames in a horizontal direction by using correlation between vertical projections of every text line in the binary images of the two video frames; and

outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

(addition 35) A computer-readable storage medium storing a program for a computer that extracts at least one text line region from a given image, characterized in that the program directs the computer to perform a process comprising:

generating edge information of the given image;

generating a binary image of candidate character strokes in the given image by using the edge information;

removing a false stroke from the binary image by using the edge information;

combining a plurality of strokes into a text line region;

removing a false character stroke from the text line region and reforming the text line region;

binarizing the text line region by using a height of the text line region; and

outputting a binary image of the text line region.

(addition 36) A computer-readable storage medium storing a program for a computer that extracts at least one text line region from a given image, characterized in that the program directs the computer to perform a process comprising:

generating an edge image of the given image;

generating a binary image of candidate character strokes in the given image by using the edge image;

checking an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels indicating an edge in the edge image;

determining that the stroke is a valid stroke if the overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold;

removing the invalid stroke; and

outputting information of remaining strokes in the binary image of the candidate character strokes.

(addition 37) A propagation signal for propagating a program to a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

removing redundant video frames from the given video frames;

removing video frames that do not contain a text area from the given video frames;

detecting and removing redundant video frames caused by image shifting from the given video frames; and

outputting remaining video frames as candidate text change frames.

(addition 38) A propagation signal for propagating a program to a computer that selects a plurality of video frames including text contents from given video frames,

characterized in that the program directs the computer to perform a process comprising:

determining whether two image blocks in the same position in two video frames of given video frames are a valid block pair that has an ability to show a change of image contents;

calculating a similarity of two image blocks of the valid block pair and determining whether the two image blocks are similar;

determining whether the two video frames are similar by using a ratio of a number of similar image blocks to a total number of valid block pairs; and

outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

(addition 39) A propagation signal for propagating a program to a computer that selects a plurality of video frames including text contents from given video frames, characterized in that the program directs the computer to perform a process comprising:

generating a first binary image of a video frame of the given video frames;

determining a position of a text line region by using a horizontal projection and a vertical projection of the first binary image;

generating a second binary image of every text line region;

determining validity of a text line region by using a difference between the first binary image and the second binary image and a fill rate of a number of foreground

pixels in the text line region to a total number of pixels in the text line region;

confirming whether a set of continuous video frames are non-text frames that do not contain a text area by using a number of valid text line regions in the set of continuous video frames; and

outputting remaining video frames after the non-text frames are removed, as candidate text change frames.

(addition 40) A propagation signal for propagating a program to a computer that selects a plurality of video frames including text contents from given videc frames, characterized in that the program directs the computer to perform a process comprising:

generating binary images of two video frames of the given video frames;

determining a vertical position of every text line region by using horizontal projections of the binary images of the two video frames;

determining a vertical offset of image shifting between the two video frames and a similarity of the two video frames in a vertical direction by using correlation between the horizontal projections;

determining a horizontal offset of the image shifting and a similarity of the two video frames in a horizontal direction by using correlation between vertical projections of every text line in the binary images of the two video frames; and

outputting remaining video frames after a similar video frame is removed, as candidate text change frames.

(addition 41)　A propagation signal for propagating a program to a computer that extracts at least one text line region from a given image, characterized in that the program directs the computer to perform a process comprising:

generating edge information of the given image;

generating a binary image of candidate character strokes in the given image by using the edge information;

removing a false stroke from the binary image by using the edge information;

combining a plurality of strokes into a text line region;

removing a false character stroke from the text line region and reforming the text line region;

binarizing the text line region by using a height of the text line region; and

outputting a binary image of the text line region.

(addition 42)　A propagation signal for propagating a program to a computer that extracts at least one text line region from a given image, characterized in that the program directs the computer to perform a process comprising:

generating an edge image of the given image;

generating a binary image of candidate character strokes in the given image by using the edge image;

checking an overlap rate of a contour of a stroke in the binary image of the candidate character strokes by pixels indicating an edge in the edge image;

determining that the stroke is a valid stroke if the

overlap rate is greater than a predefined threshold and an invalid stroke if the overlap rate is less than the predefined threshold;

removing the invalid stroke; and

outputting information of remaining strokes in the binary image of the candidate character strokes.

(addition 43) A text change frame detection method for selecting a plurality of video frames that includes text contents from given video frames, said method comprising:

removing redundant video frames from the given video frames;

removing video frames that do not contain a text area from the given video frames;

detecting and removing redundant video frames caused by image shifting from the given video frames; and

presenting remaining video frames as candidate text change frames.

(addition 44) A text extraction method for extracting at least one text line region from a given image, said method comprising:

generating edge information of the given image;

generating a binary image of candidate character strokes in the given image by using the edge information;

removing a false stroke from the binary image by using the edge information;

combining a plurality of strokes into a text line region;

removing a false character stroke from the text line region and reforming the text line region;

binarizing the text line region by using a height of the text line region; and

presenting a binary image of the text line region.

Effect of Invention

According to the present invention, duplicate video frames, shifting video frames as well as video frames that do not contain a text area can be removed in a very fast speed from given video frames. Further, the text line region in a video frame can be accurately binarized since the false strokes are detected and removed as much as possible.

【図面の簡単な説明】

Fig. 1

the drawing which shows the configuration of the video text processing apparatus according to the present invention

Fig. 2

the processing flowchart of the video text processing apparatus

Fig. 3

the block diagram which shows the configuration of the text change frame detection apparatus according to the present invention

Fig. 4

the drawing which shows the configuration of the frame similarity measurement unit

Fig. 5

the drawing which shows the configuration of the text frame detection and verification unit

Fig. 6

the drawing which shows the configuration of the image shifting detection unit

Fig. 7

the drawing which shows the first frame that has a text content

Fig. 8

the drawing which shows the second frame that has a text content

Fig. 9

the drawing which shows the processing result of the frame similarity measurement unit

Fig. 10

the flowchart of the operation of the frame similarity measurement unit

Fig. 11

the flowchart of determination of the similarity of two frames

Fig. 12

the flowchart of the operation of the image block validation unit

Fig. 13

the flowchart of the operation of the image block similarity measurement unit

Fig. 14

the drawing which shows the original video frame for text frame detection and verification

Fig. 15

the drawing which shows the first binary image resulted from fast and simple image binarization

Fig. 16

the drawing which shows the result of horizontal projection

Fig. 17

the drawing which shows the result of projection regularization

Fig. 18

the drawing which shows the result of vertical binary projection in every candidate text line

Fig. 19

the drawing which shows the result of text line region determination

Fig. 20

the drawing which shows two pairs of binary images for two candidate text line regions

Fig. 21

the drawing which shows detected text line regions

Fig. 22

the flowchart of the operation of the text frame detection and verification unit (No. 1)

Fig. 23

the flowchart of the operation of the text frame detection and verification unit (No. 2)

Fig. 24

the flowchart of the operation of the fast and simple image binarization unit

Fig. 25

the flowchart of Niblack's image binarization method

Fig. 26

the flowchart of the operation of the text line region determination unit

Fig. 27

the flowchart of horizontal image projection

Fig. 28

the flowchart of projection smoothing

Fig. 29

the flowchart of projection regularization

Fig. 30

the drawing which shows examples of the max and min in a projection

Fig. 31

the flowchart of the operation of the text line confirmation unit

Fig. 32

the flowchart of the operation of the image shifting detection unit (No. 1)

Fig. 33

the flowchart of the operation of the image shifting detection unit (No. 2)

Fig. 34

the drawing which shows the configuration of the text extraction apparatus according to the present invention

Fig. 35

the drawing which shows the configuration of the edge image generation unit

Fig. 36

the drawing which shows the configuration of the stroke image generation unit

Fig. 37

the drawing which shows the configuration of the stroke filtering unit

Fig. 38

the drawing which shows the configuration of the text line region formation unit

Fig. 39

the drawing which shows the configuration of the text line verification unit

Fig. 40

the drawing which shows the configuration of the text line binarization unit

Fig. 41

the drawing which shows the original video frame for text extraction

Fig. 42

the drawing which shows the result of edge image generation

Fig. 43

the drawing which shows the result of stroke generation

Fig. 44

the drawing which shows the result of stroke filtering

Fig. 45

the drawing which shows the result of text line region formation

Fig. 46

the drawing which shows the final binarized text line regions

Fig. 47

the flowchart of the operation of the edge image generation unit (No. 1)

Fig. 48

the flowchart of the operation of the edge image generation unit (No. 2)

Fig. 49

the drawing which shows the arrangement of the neighborhood of pixel i

Fig. 50

the flowchart of the operation of the edge strength calculation unit

Fig. 51

the flowchart of the operation of the stroke image generation unit

Fig. 52

the flowchart of the operation of the stroke filtering unit

Fig. 53

the flowchart of the operation of the stroke edge coverage validation unit

Fig. 54

the flowchart of the operation of the text line region formation unit

Fig. 55

the flowchart of the operation of the stroke connection checking unit

Fig. 56

the flowchart of the operation of the text line verification unit

Fig. 57

the flowchart of the operation of the vertical false stroke detection unit

Fig. 58

the flowchart of multiple text line detection

Fig. 59

the flowchart of the operation of the horizontal false stroke detection unit

Fig. 60

the drawing which shows the first false stroke

Fig. 61

the drawing which shows the second false stroke

Fig. 62

the flowchart of the operation of the text line binarization unit

Fig. 63

the drawing which shows the configuration of an information processing apparatus

Fig. 64

the drawing which shows storage media

Description of Notations

101　video data

102　TV video camera

103　video decomposition unit

104　text change frame detection apparatus

105　text extraction apparatus

106, 1803　database

301　frame similarity measurement unit

302　text frame detection and verification unit

303　image shifting detection unit

311　image block validation unit

312　image block similarity measurement unit

313　frame similarity judgment unit

321, 331　fast and simple image binarization unit

322　text line region determination unit

323　rebinarization unit

324　text line confirmation unit

325　text frame verification unit

332　text line vertical position determination unit

333　vertical shifting detection unit

334　horizontal shifting detection unit

901　edge image generation unit

902　stroke image generation unit

903　stroke filtering unit

904　text line region formation unit

905　text line verification unit

906　text line binarization unit

911　edge strength calculation unit

912　first edge image generation unit

913　second edge image generation unit

921　local image binarization unit

931　stroke edge coverage validation unit

932　long straight line detection unit

941　stroke connection checking unit

951　vertical false stroke detection unit

952　horizontal false stroke detection unit

953　text line reformation unit

961　automatic size calculation unit

962　block image binarization unit

1541, 1542　false stroke

1701　CPU

1702　memory

1703　input device

1704　output device

1705　external storage device

1706　medium drive device

1707　network connection device

1708　video input device

1709　bus

1710　portable storage medium

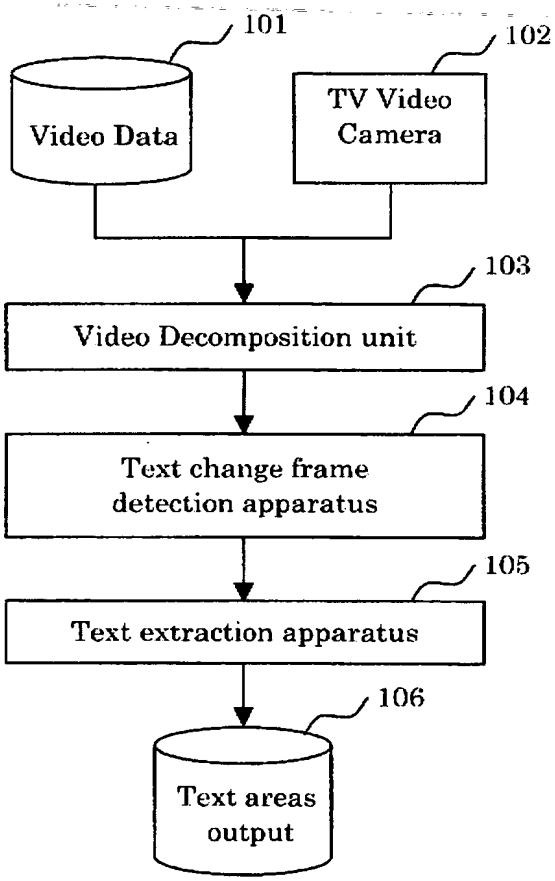
1801　server

1802　information processing apparatus

【書類名】　外国語図面

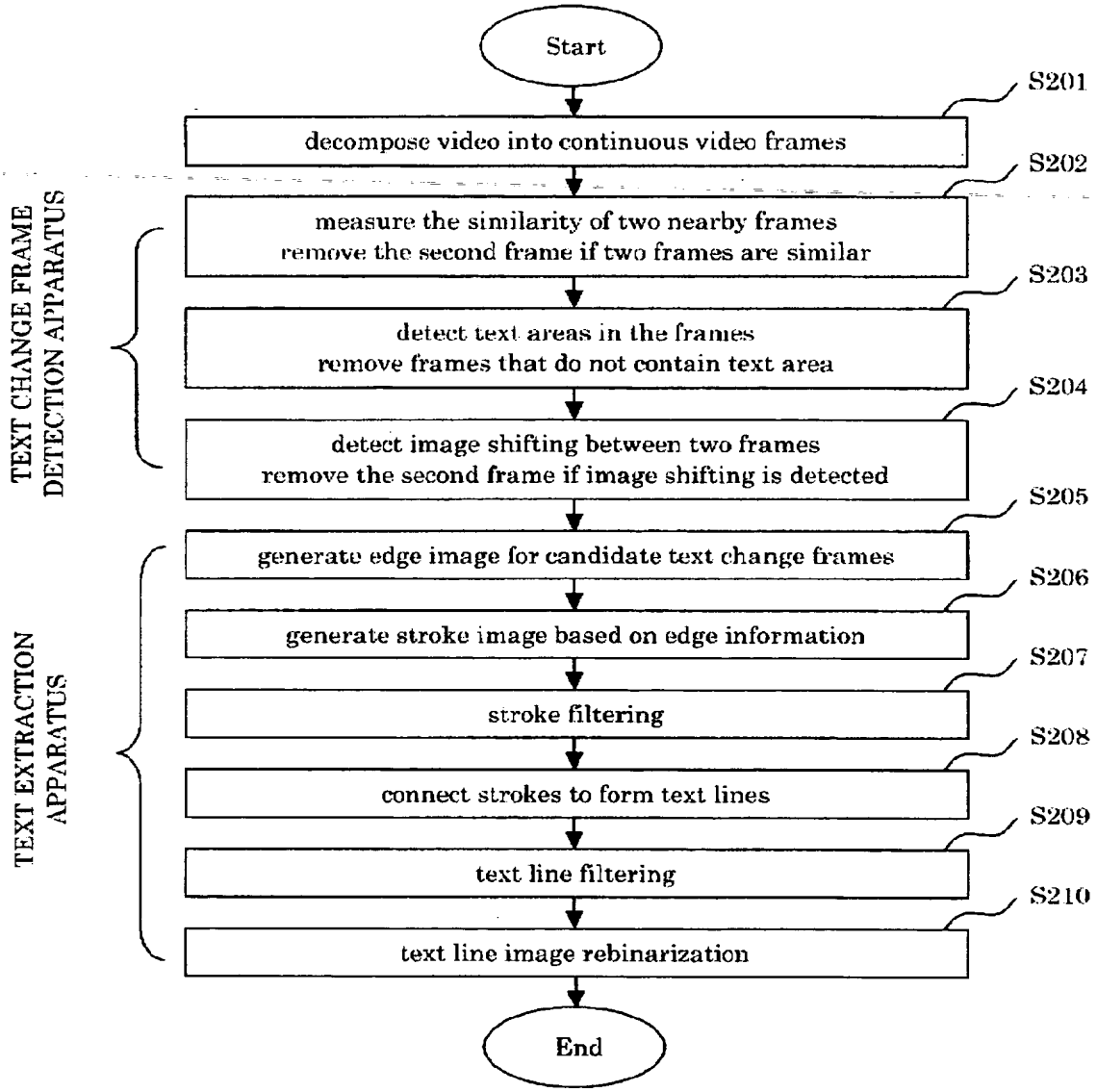
【図１】

the drawing which shows the configuration of the video text processing apparatus according to the present invention



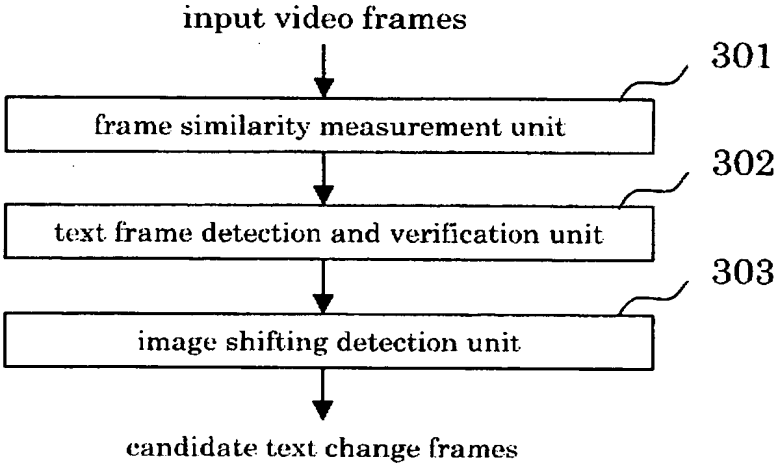
【図２】

the processing flowchart
of the video text processing apparatus



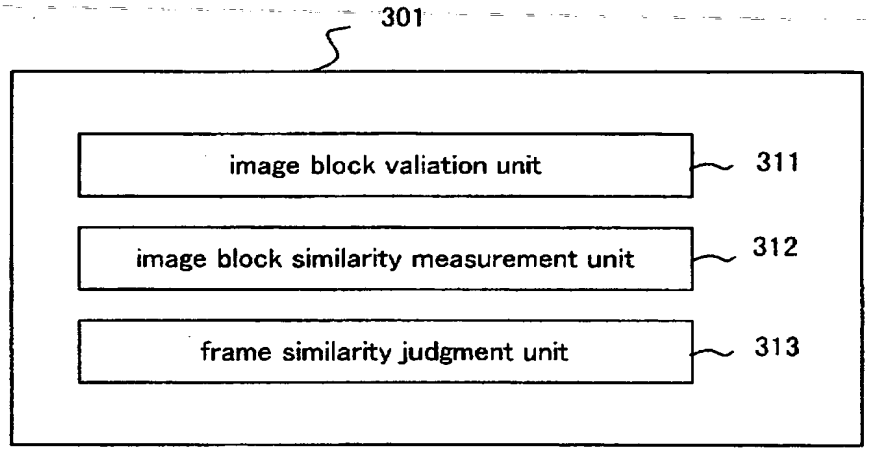
【図３】

the block diagram which shows the configuration of the text change frame detection apparatus according to the present invention



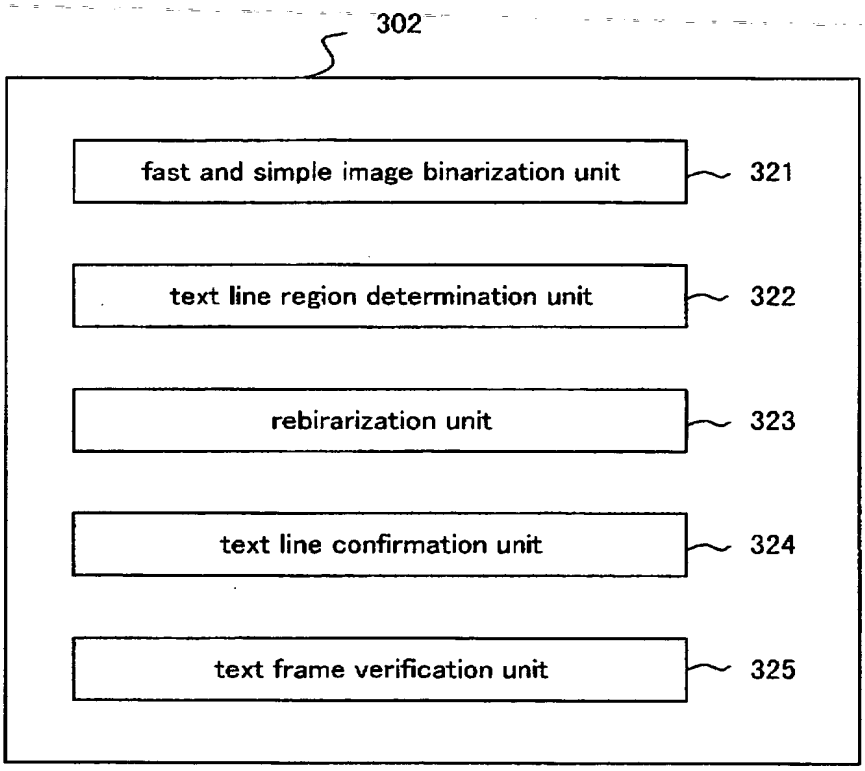
【図４】

the drawing which shows the configuration of the frame similarity measurement unit



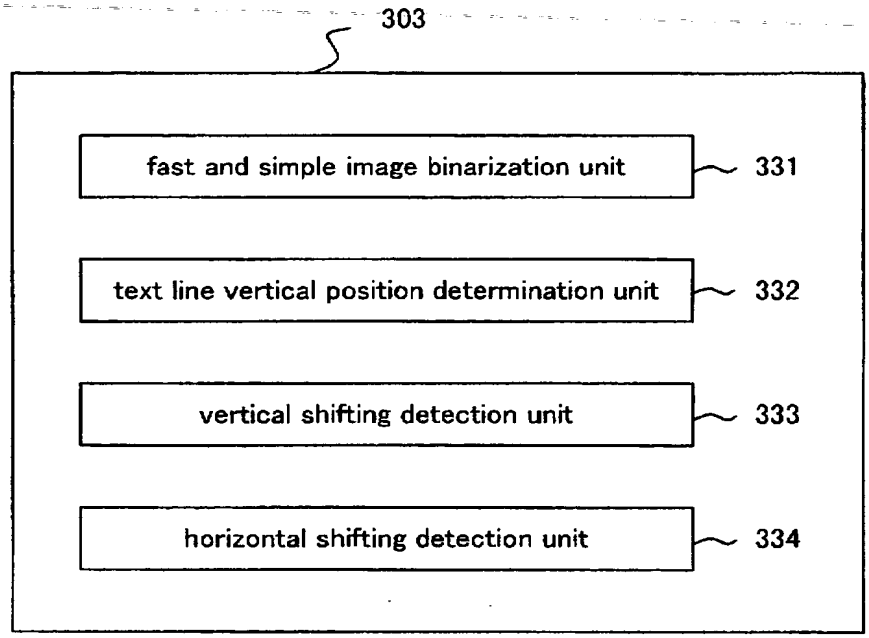
【図５】

the drawing which shows the configuration of the text frame detection and verification unit



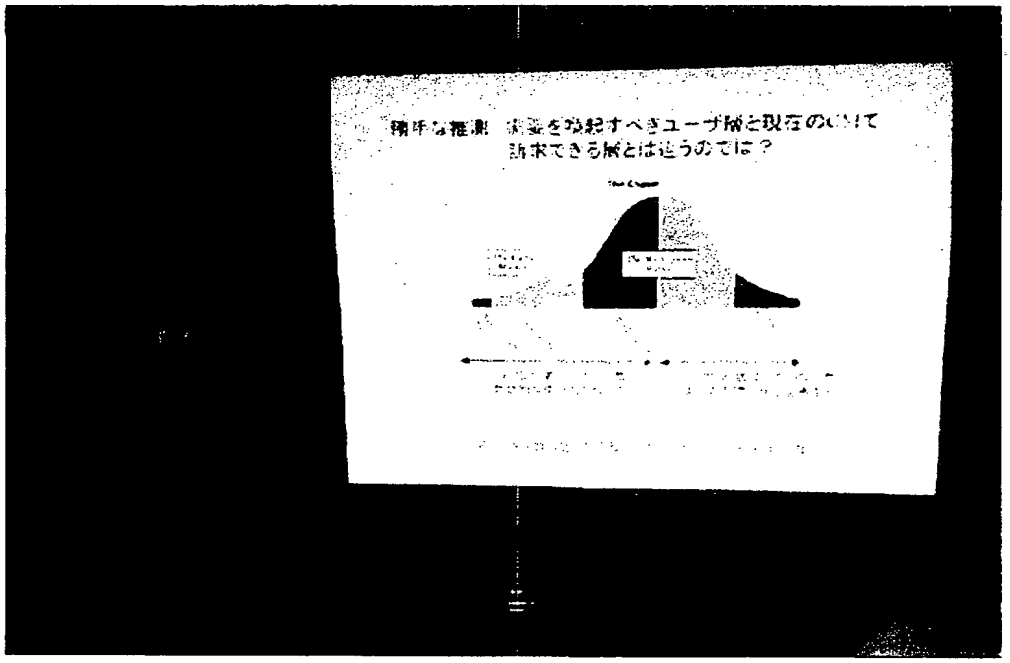
【図６】

the drawing which shows the configuration of the image shifting detection unit



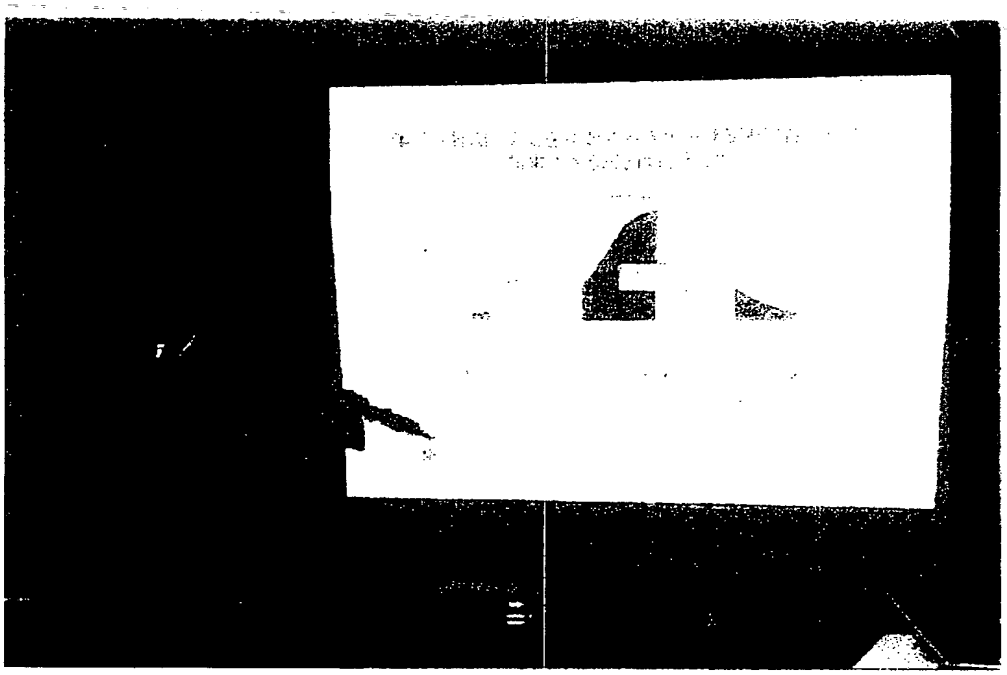
【図７】

the drawing which shows the first frame
that has a text content



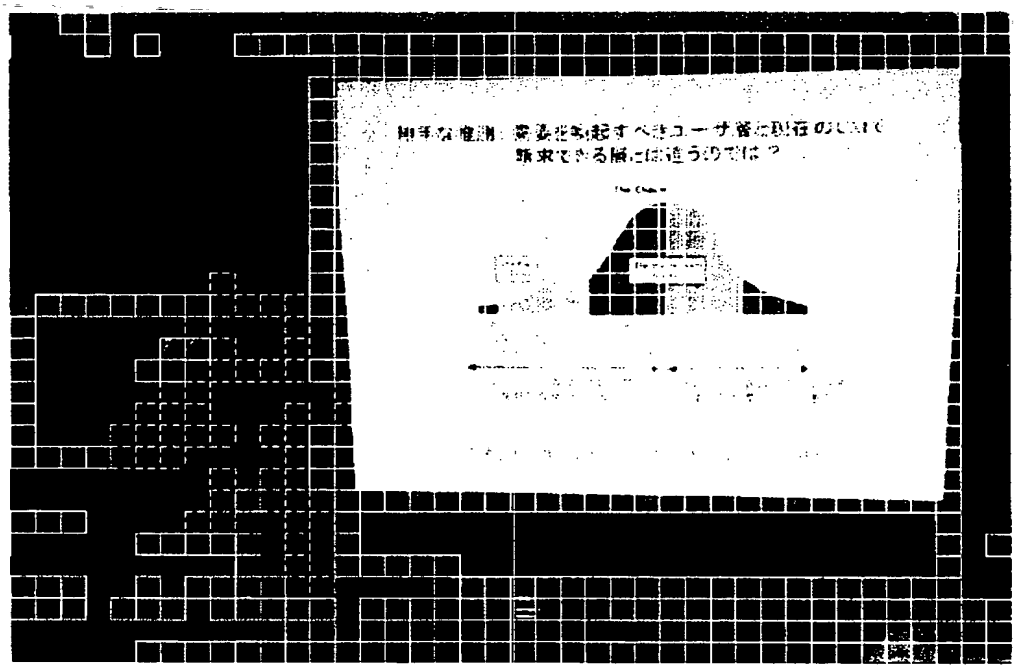
【図８】

the drawing which shows the second frame
that has a text content



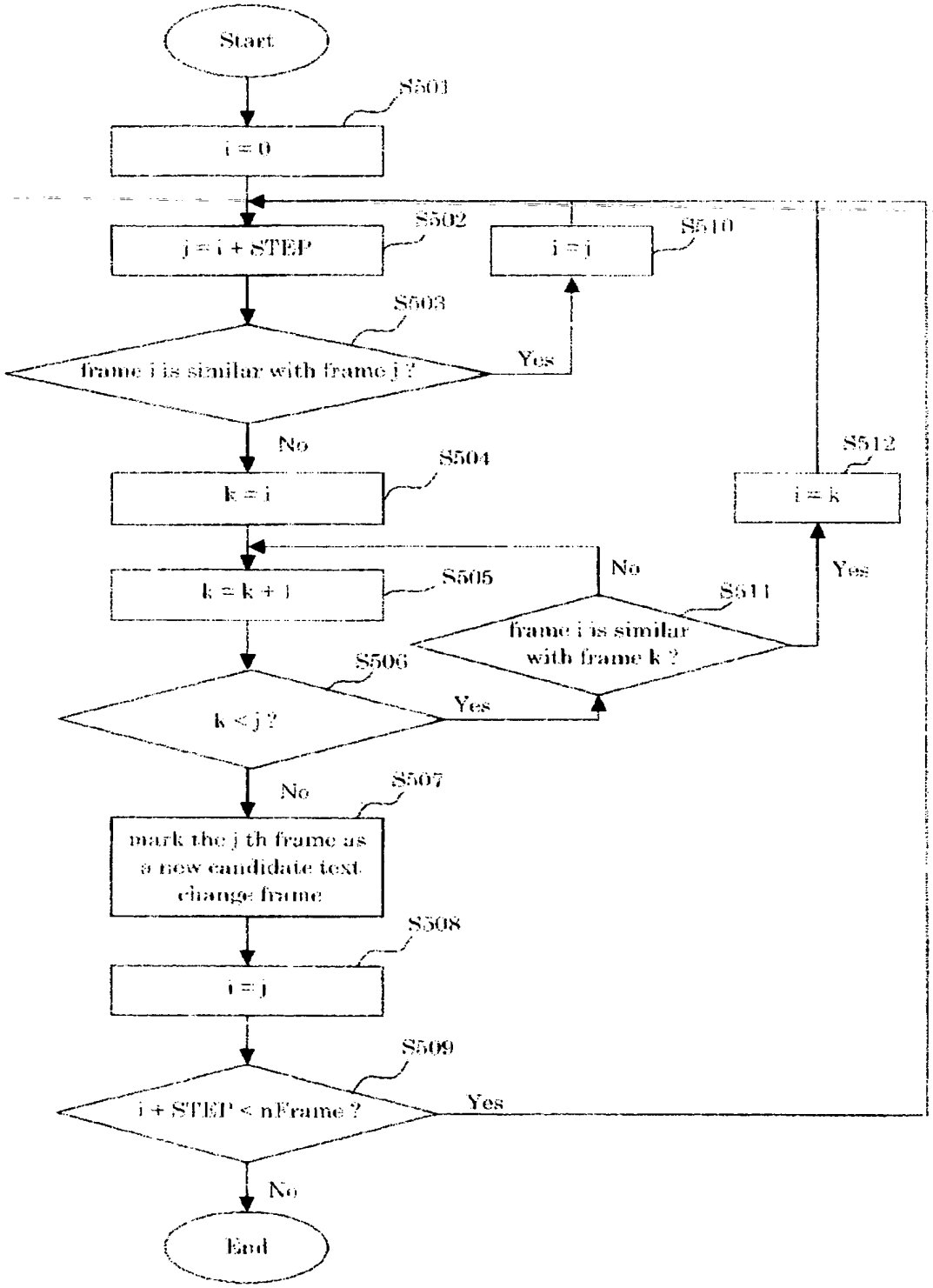
【図９】

the drawing which shows the processing result of the frame similarity measurement unit



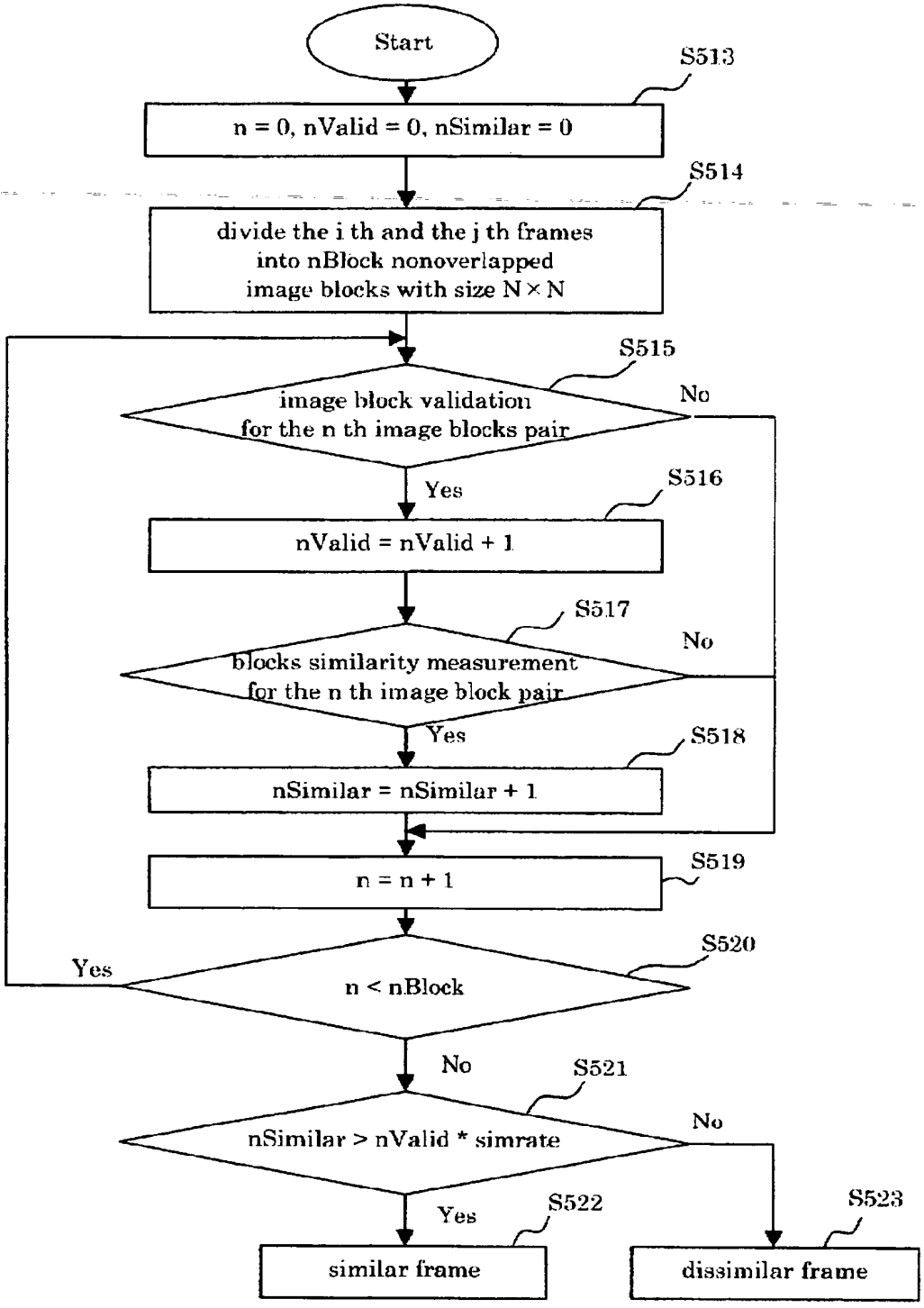
【図１０】

the flowchart of the operation
of the frame similarity measurement unit



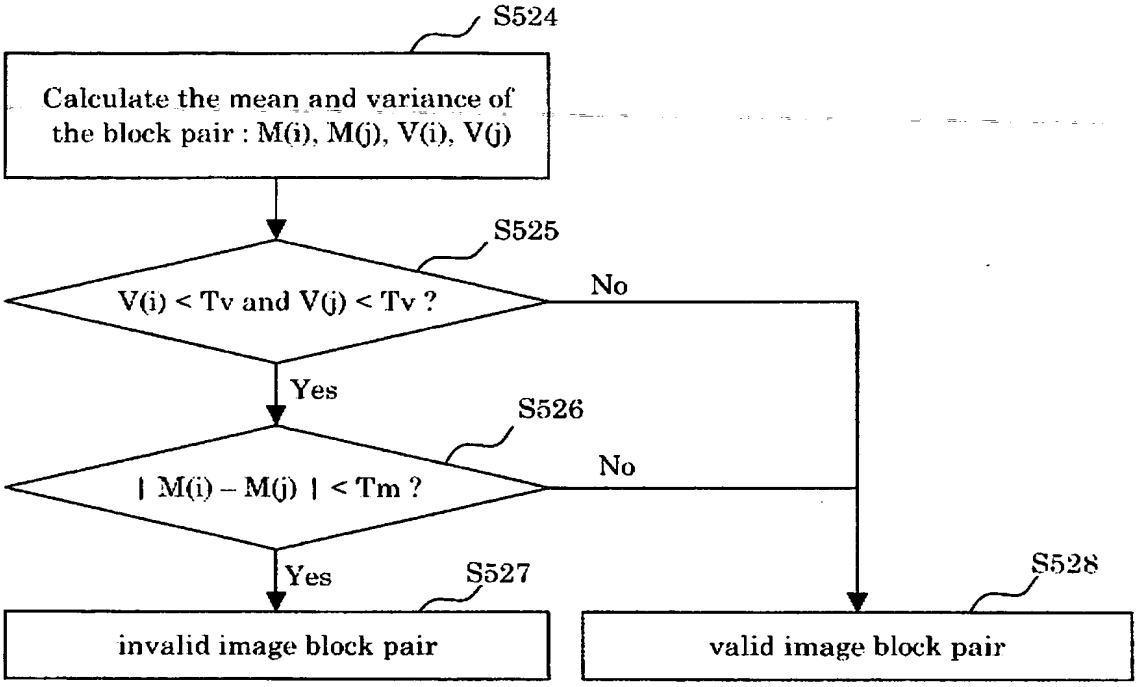
【図１１】

the flowchart of determination
of the similarity of two frames



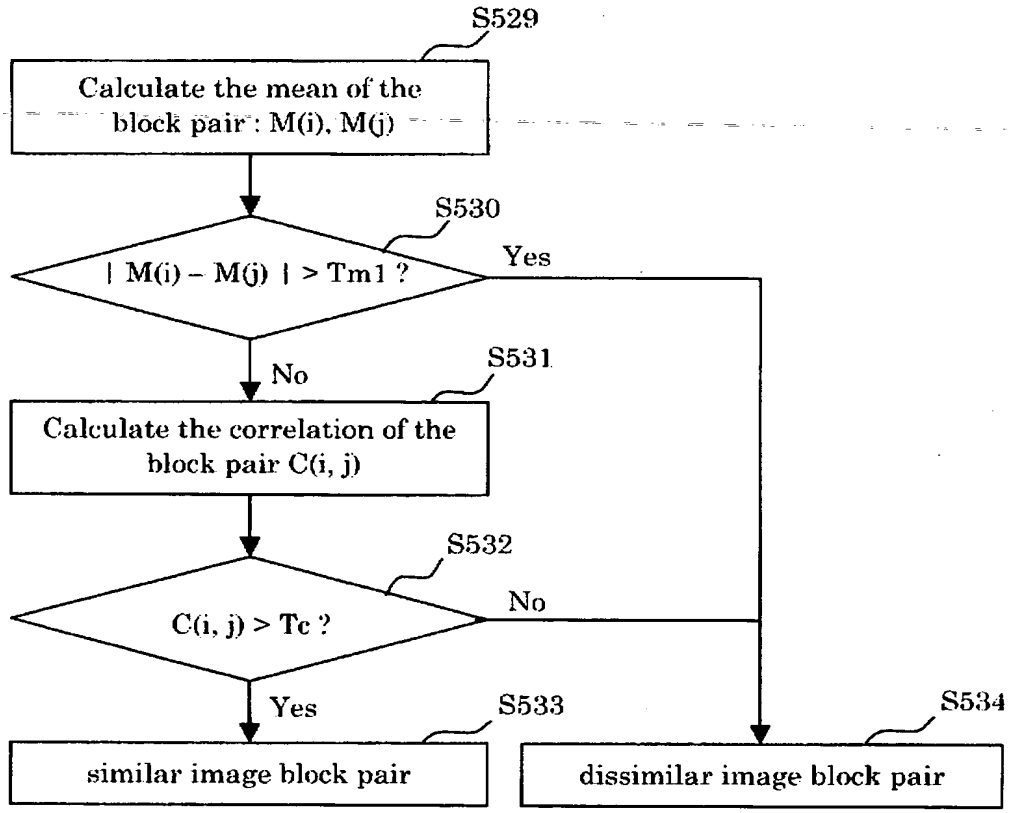
【図１２】

the flowchart of the operation
of the image block validation unit



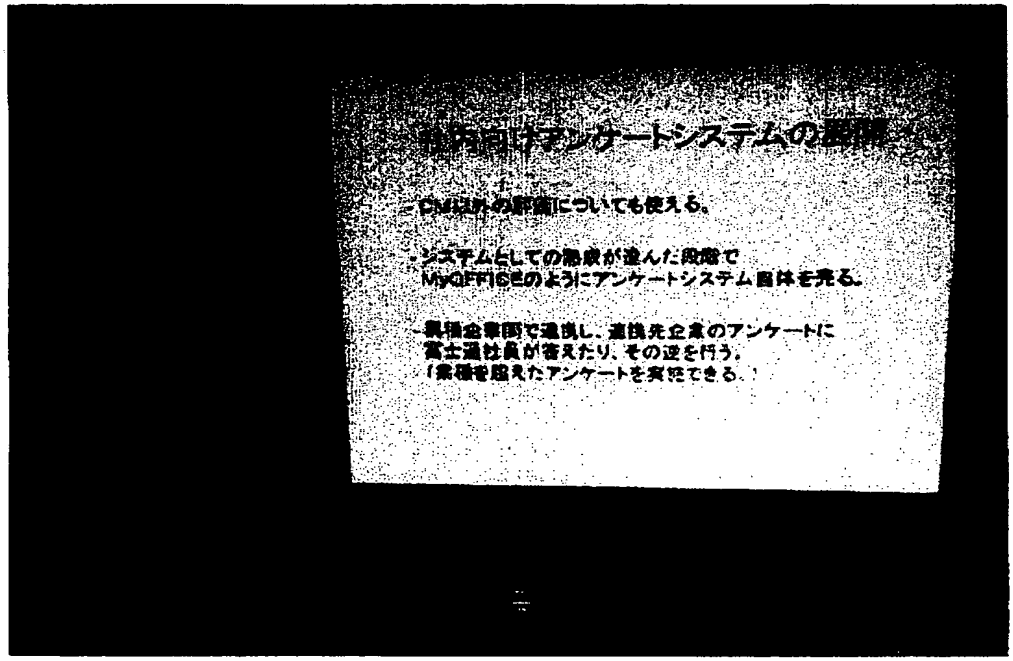
【図１３】

the flowchart of the operation
of the image block similarity measurement unit



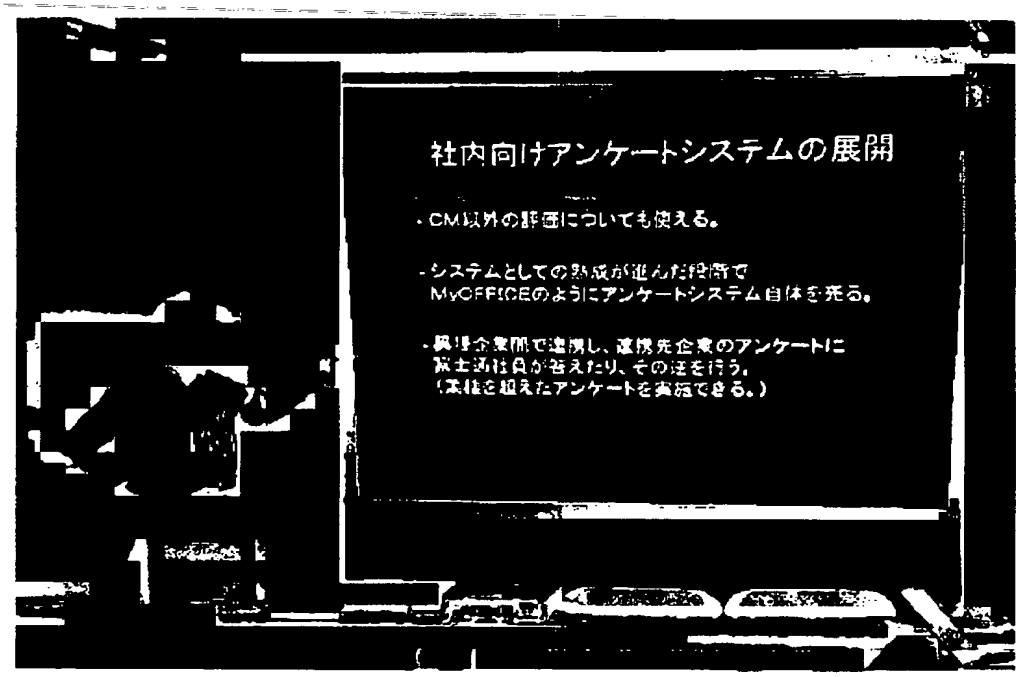
【図１４】

the drawing which shows the original video frame
for text frame detection and verification



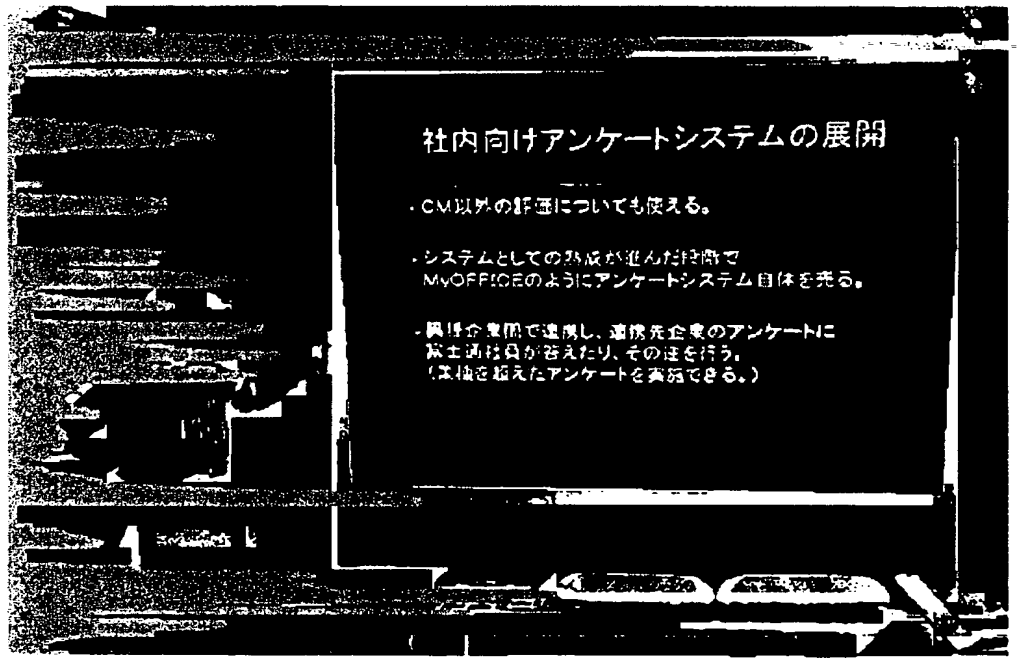
【図１５】

the drawing which shows the first binary image resulted from fast and simple image binarization



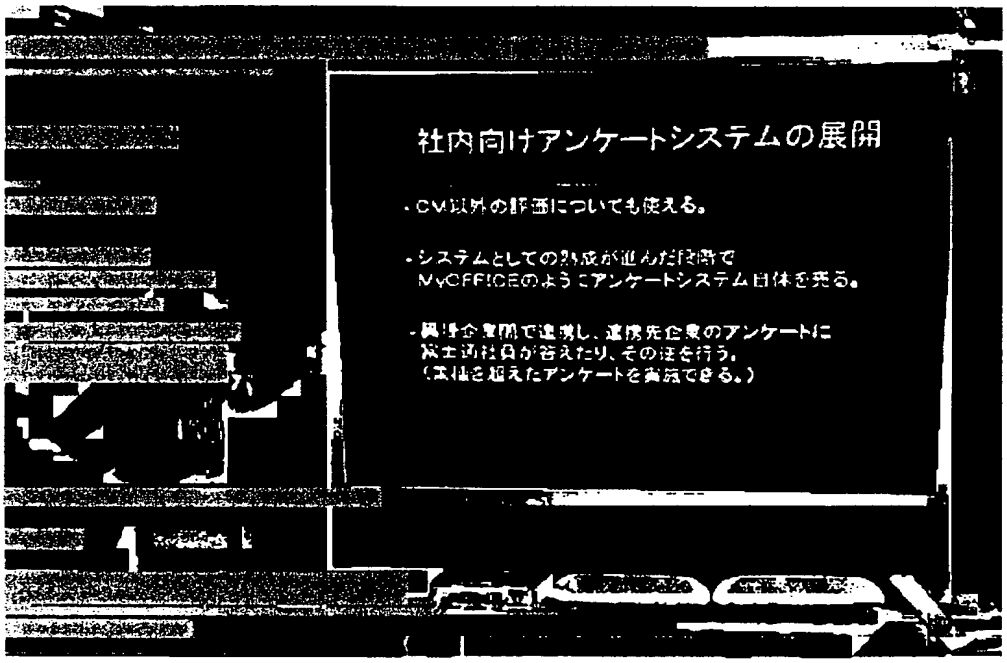
【図１６】

the drawing which shows
the result of horizontal projection



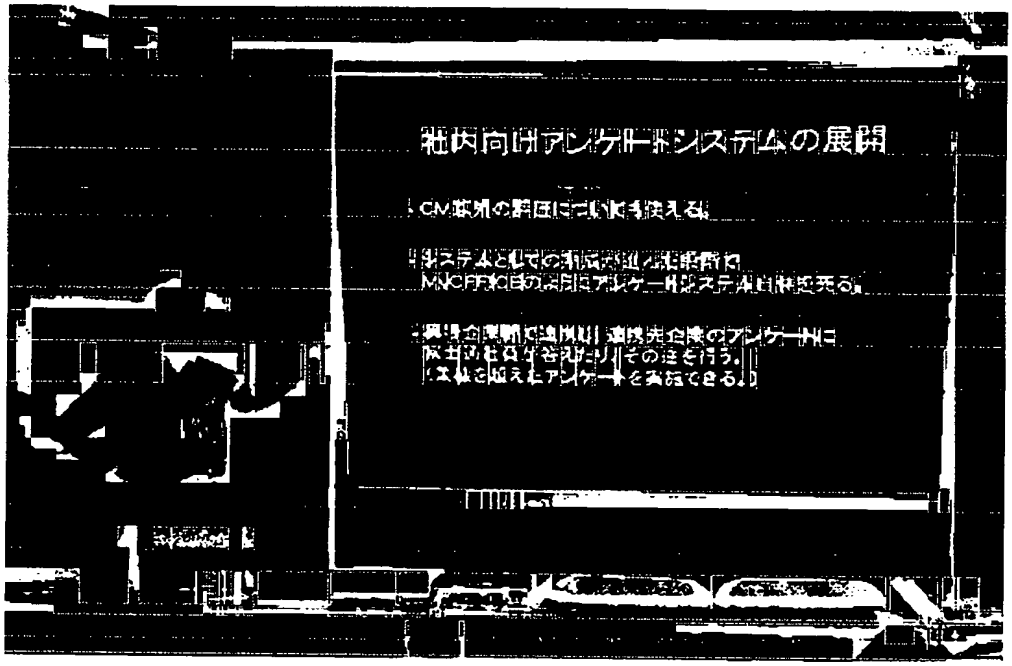
【図１７】

the drawing which shows
the result of projection regularization



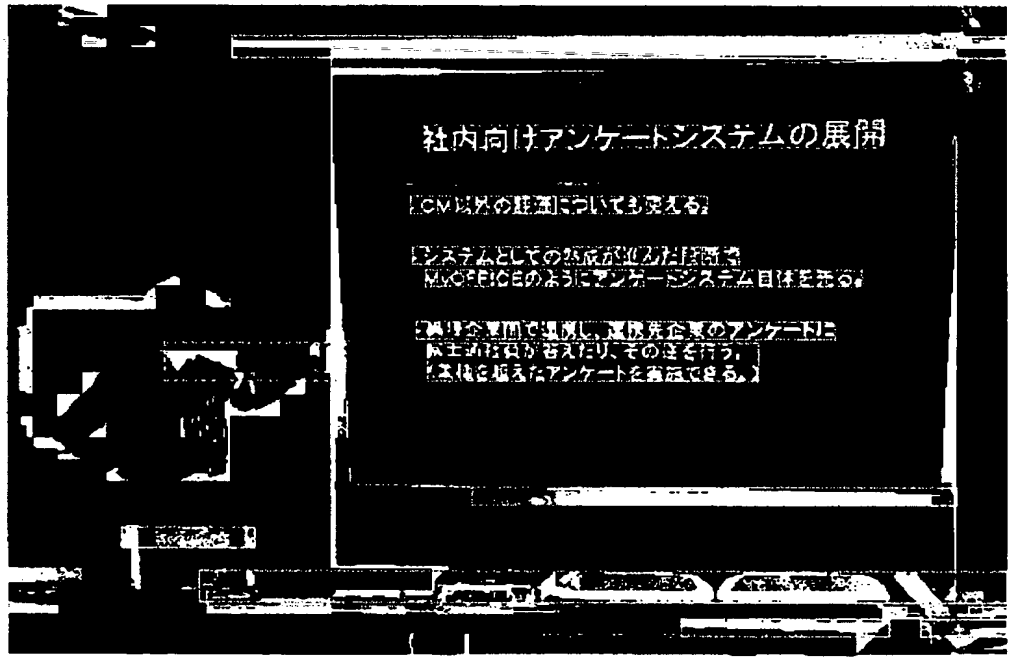
【図１８】

the drawing which shows
the result of vertical binary projection
in every candidate text line



【図１９】

the drawing which shows the result
of text line region determination

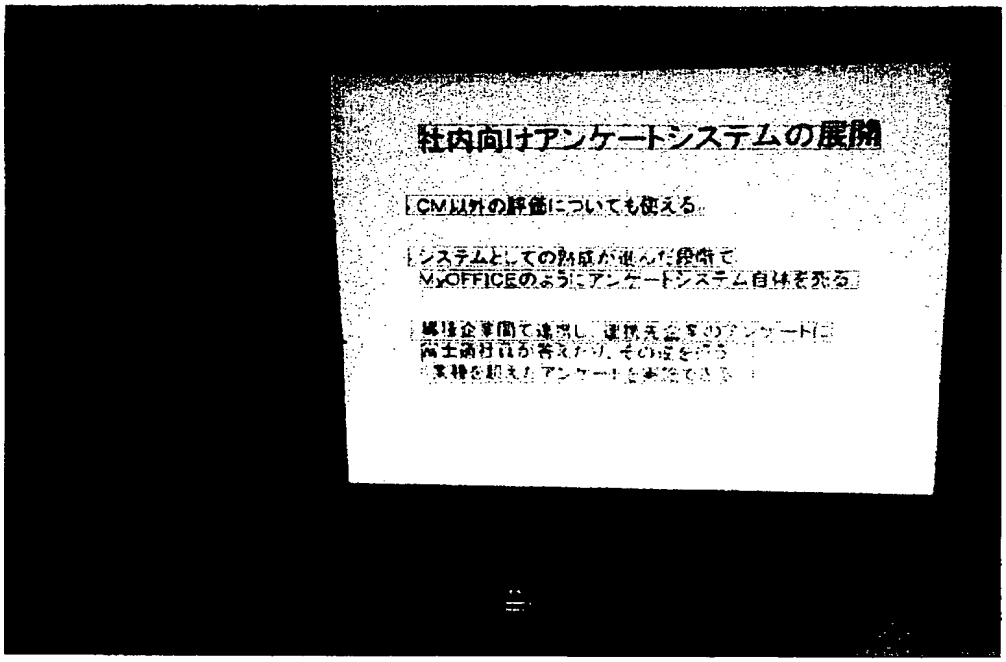


【図２０】

the drawing which shows two pairs of binary images
for two candidate text line regions

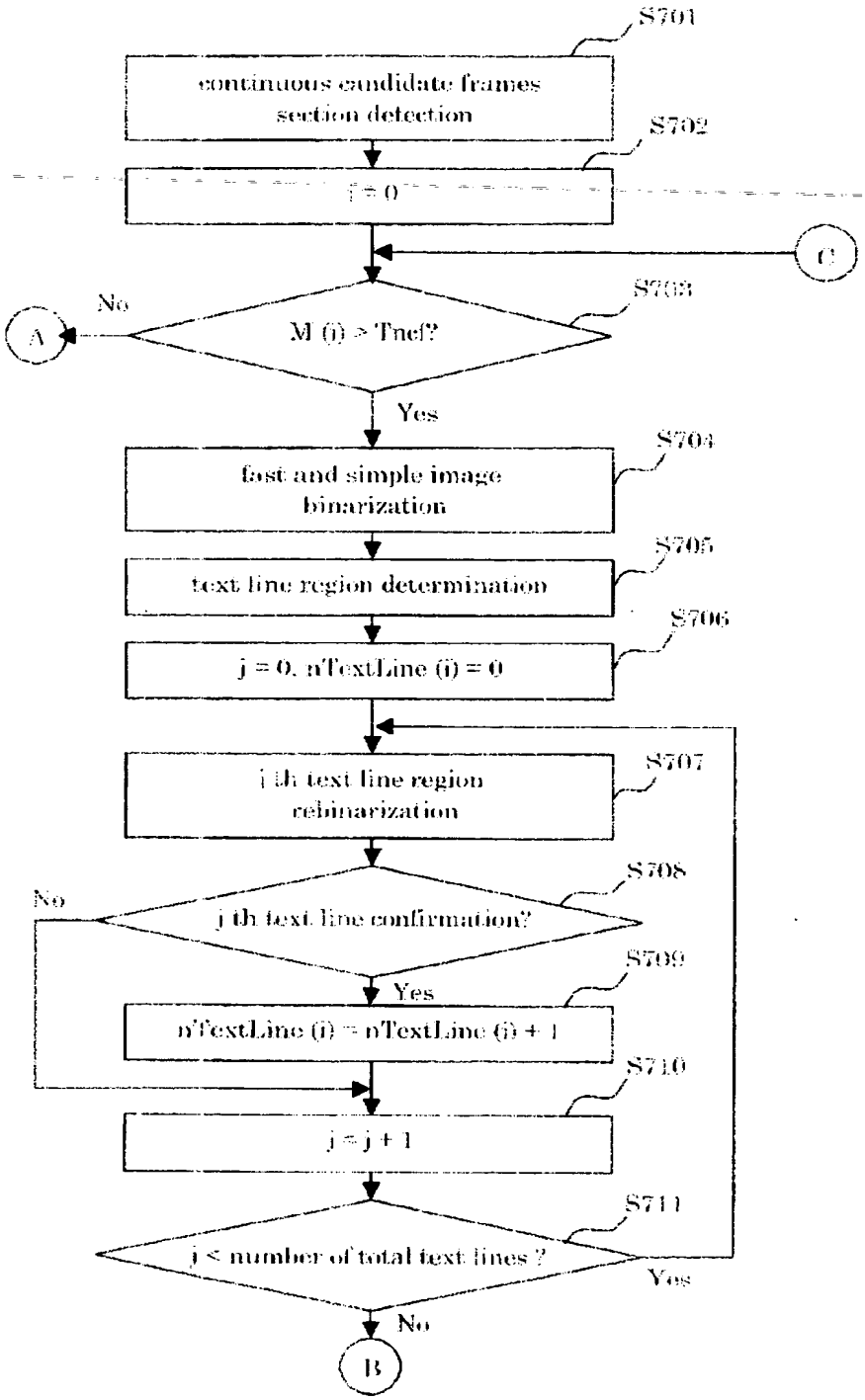
【図２１】

the drawing which shows detected text line regions



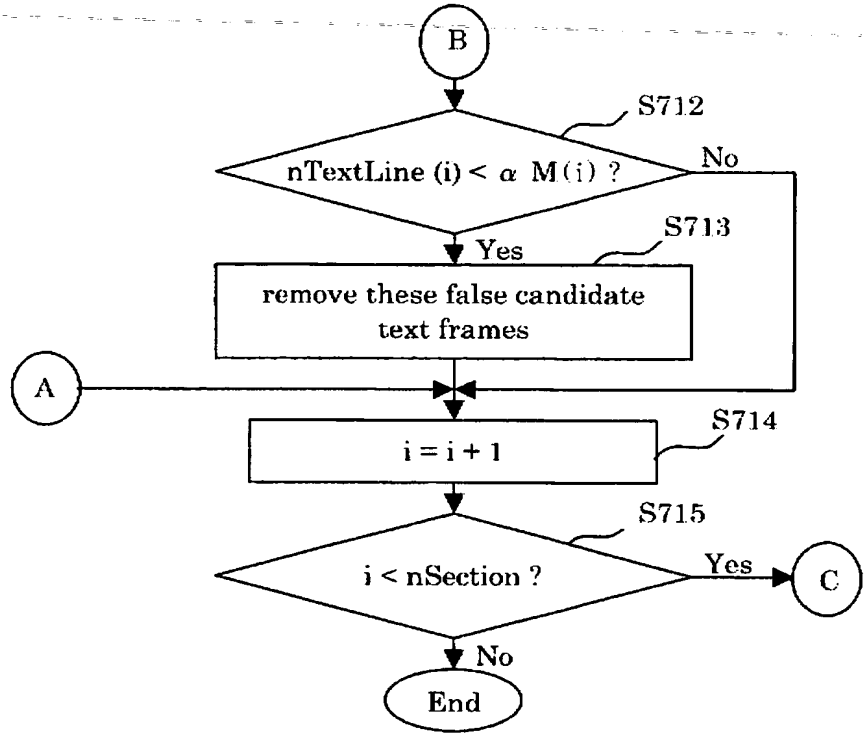
【図２２】

the flowchart of the operation of the text frame detection and verification unit (No. 1)



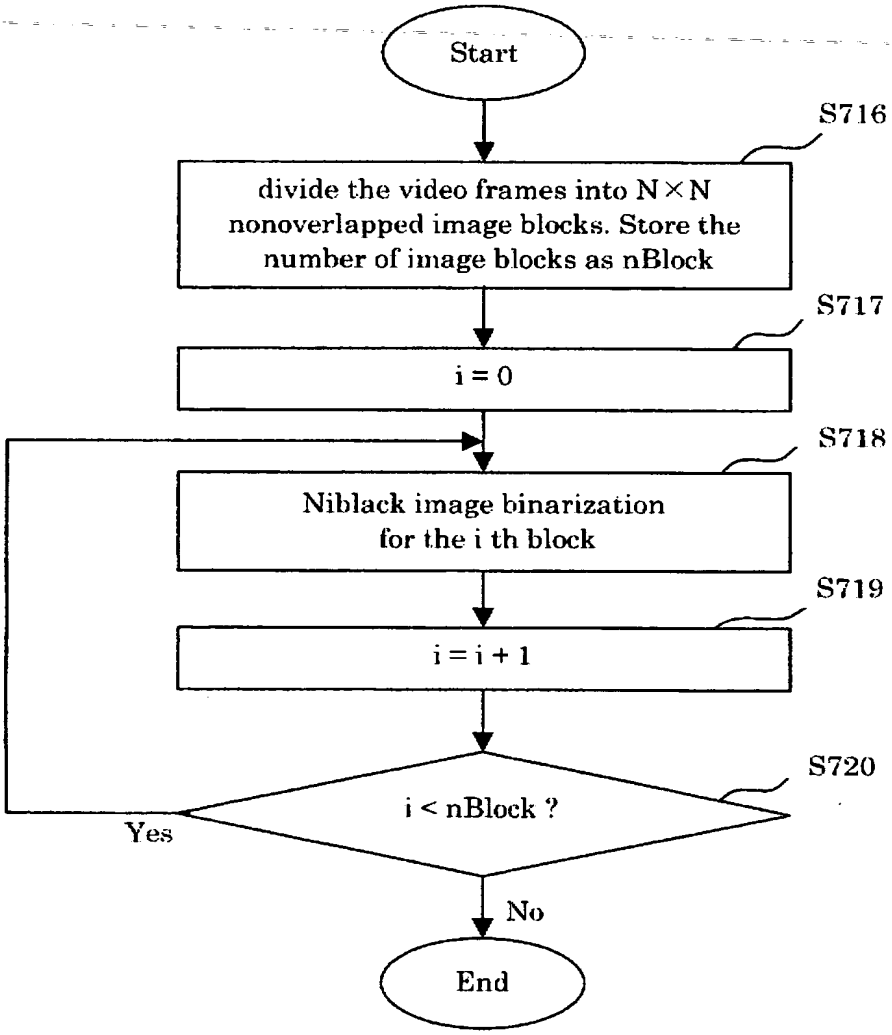
【図２３】

the flowchart of the operation of the text frame detection and verification unit (No. 2)



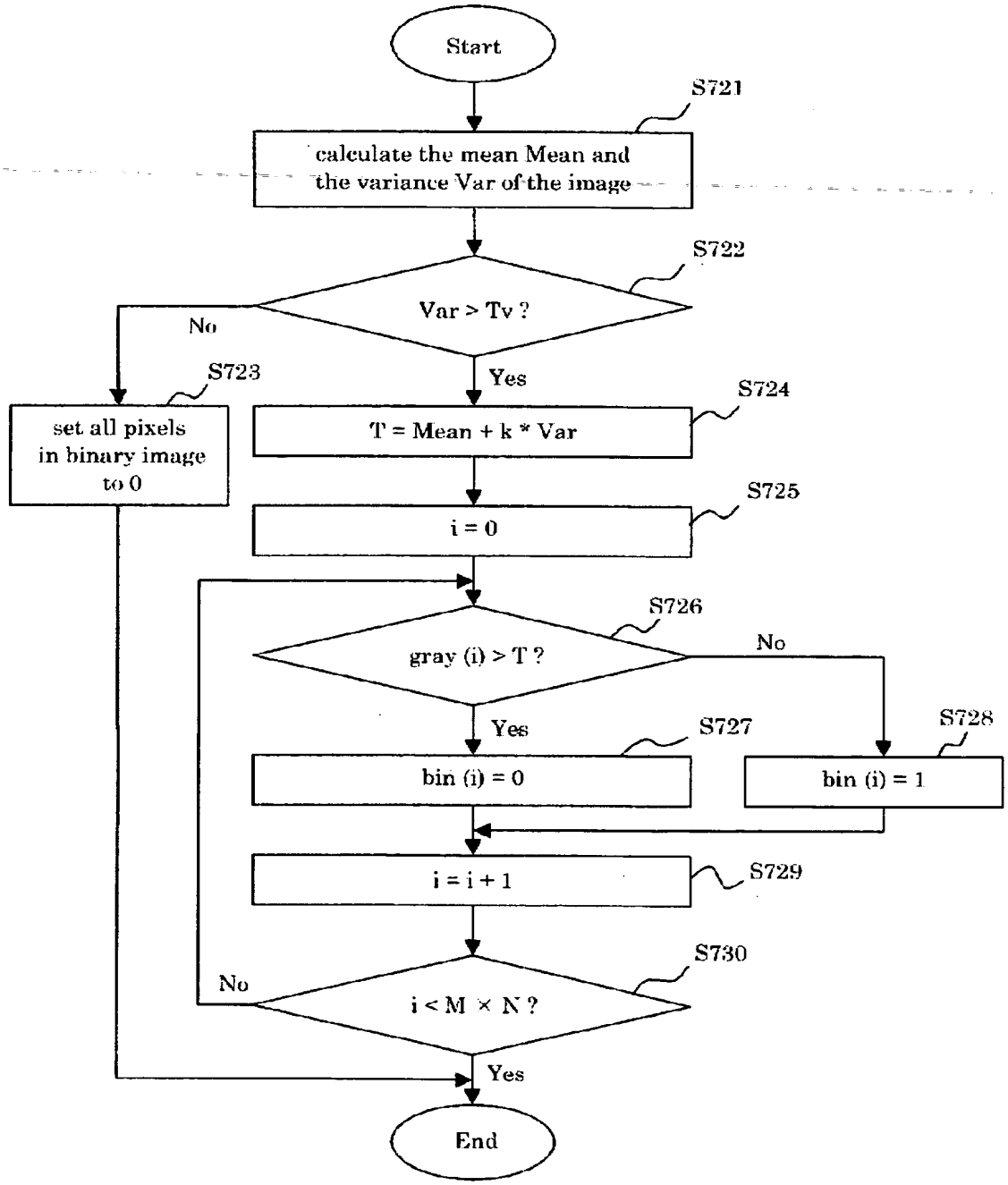
【図２４】

the flowchart of the operation
of the fast and simple image binarization unit



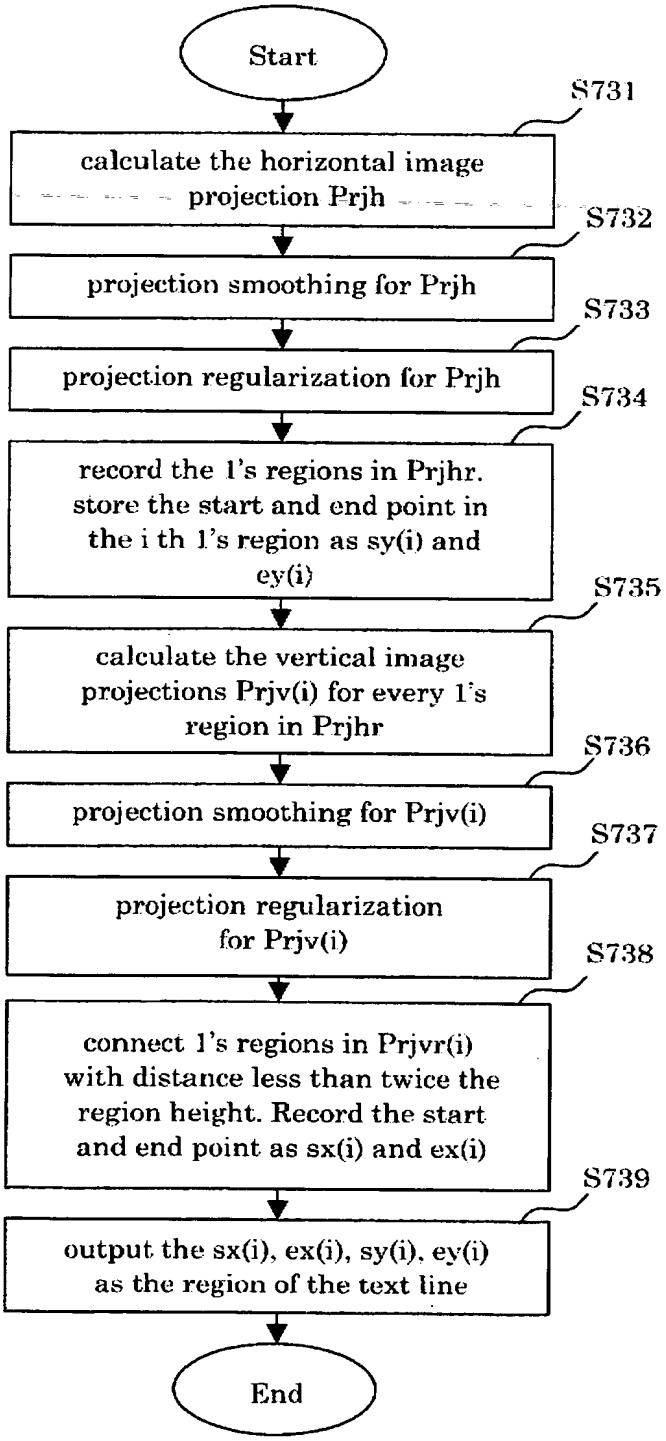
【図２５】

the flowchart of Niblack's image binarization method



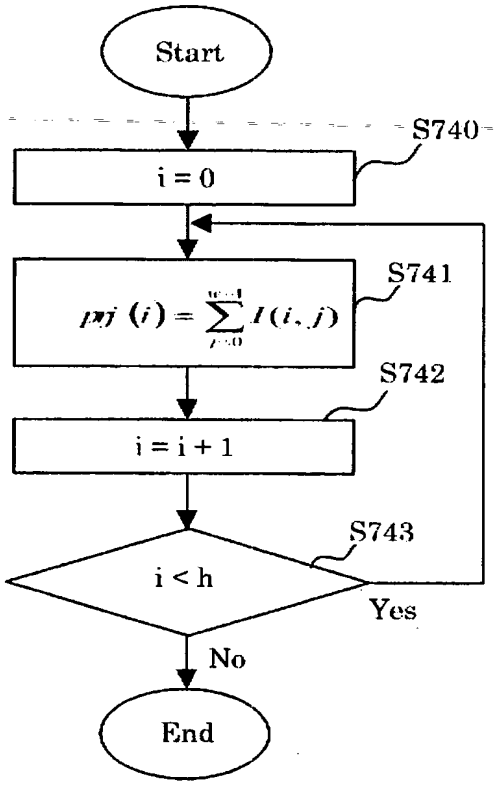
【図２６】

the flowchart of the operation
of the text line region determination unit



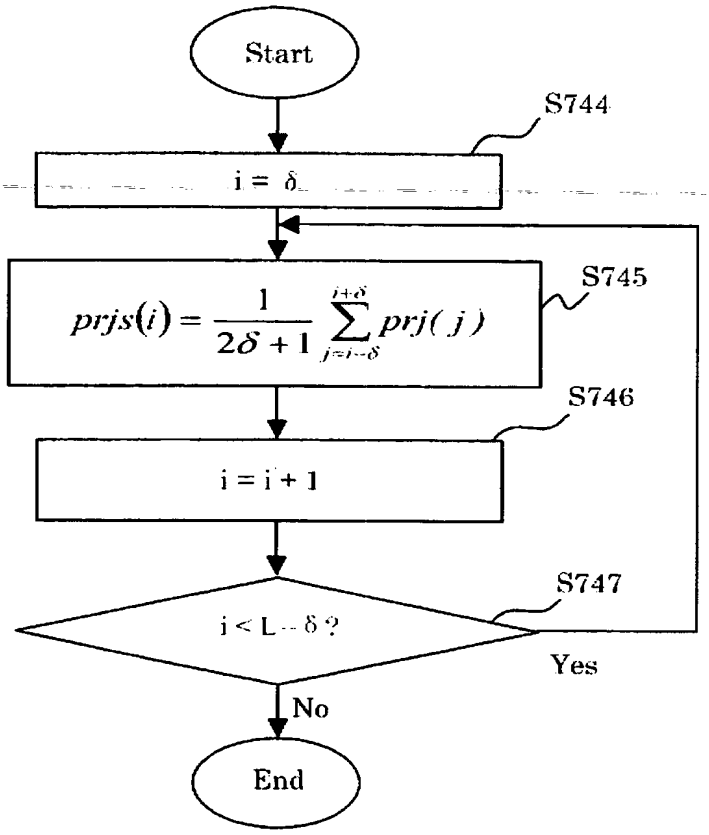
【図２７】

the flowchart of horizontal image projection



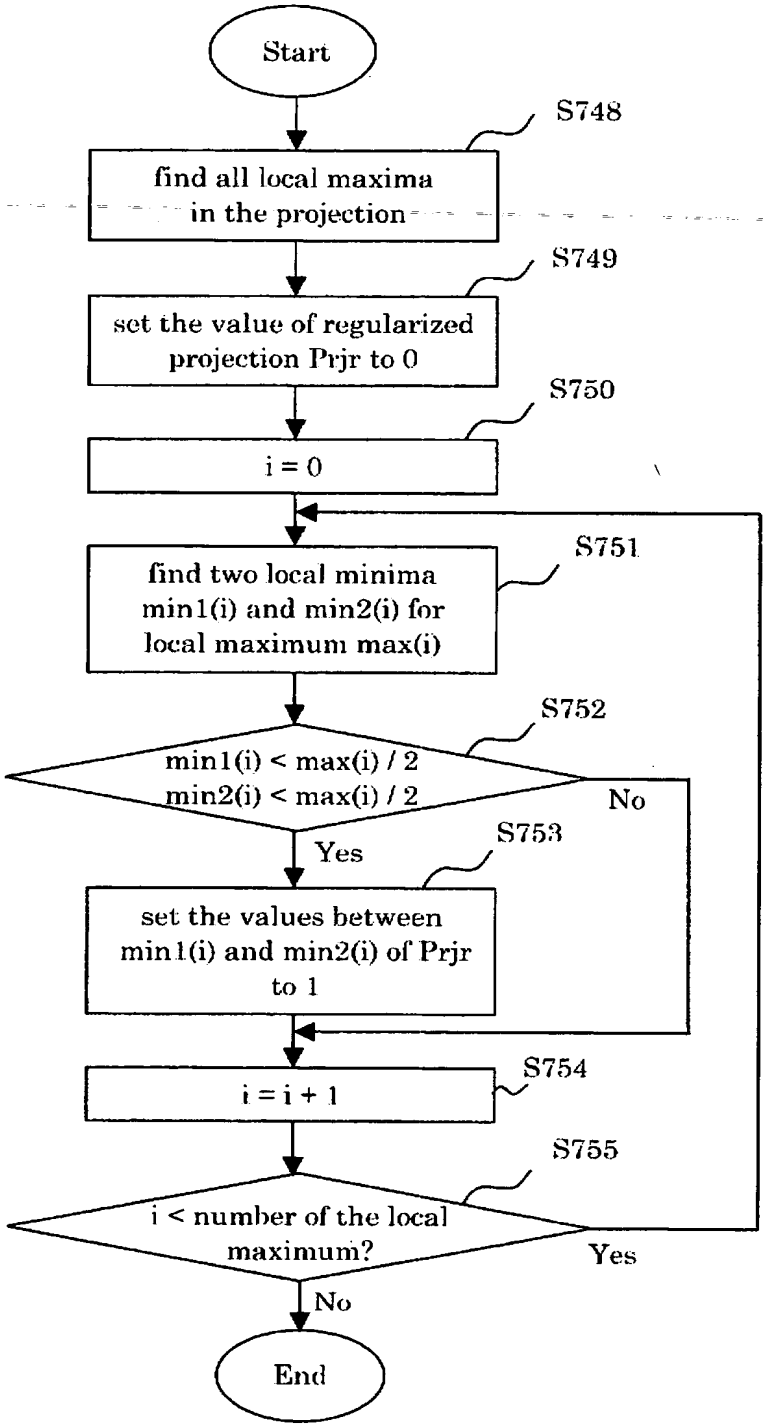
【図２８】

the flowchart of projection smoothing



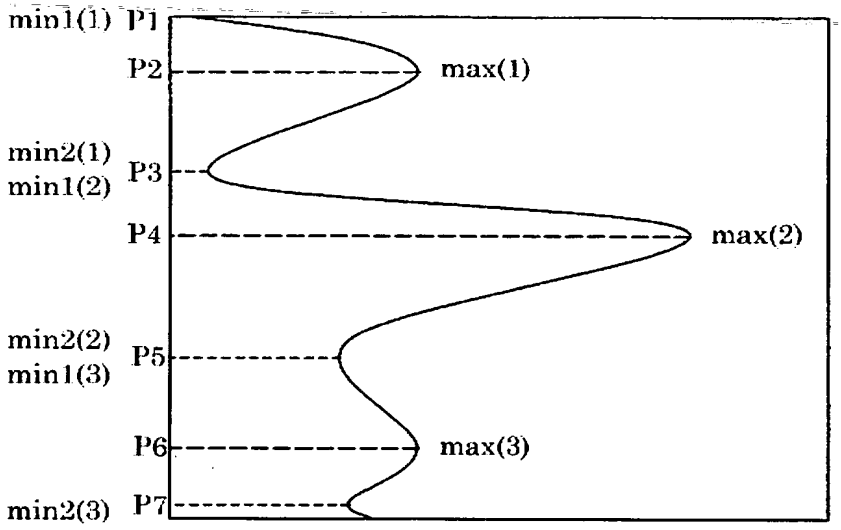
【図２９】

the flowchart of projection regularization



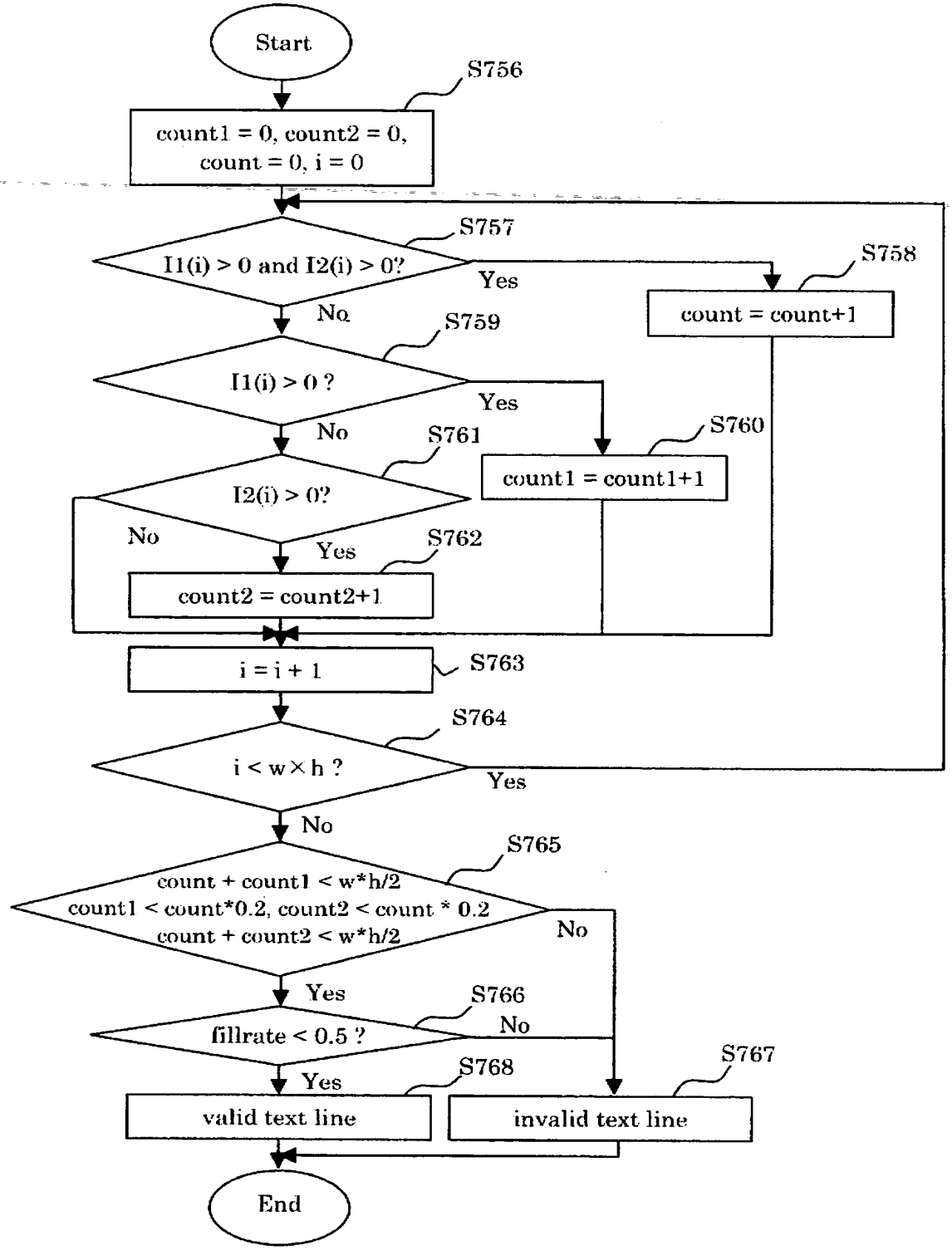
【図３０】

the drawing which shows examples of the max and min in a projection



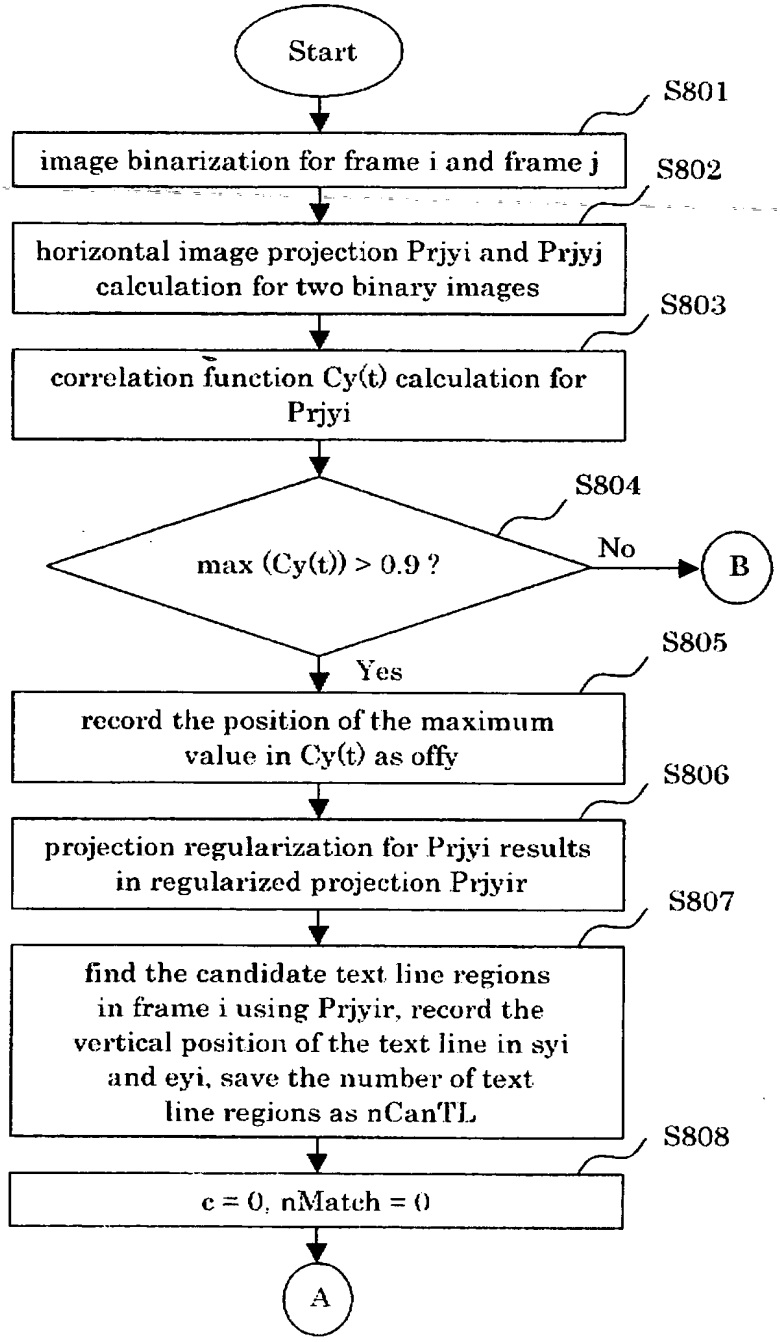
【図３１】

the flowchart of the operation
of the text line confirmation unit



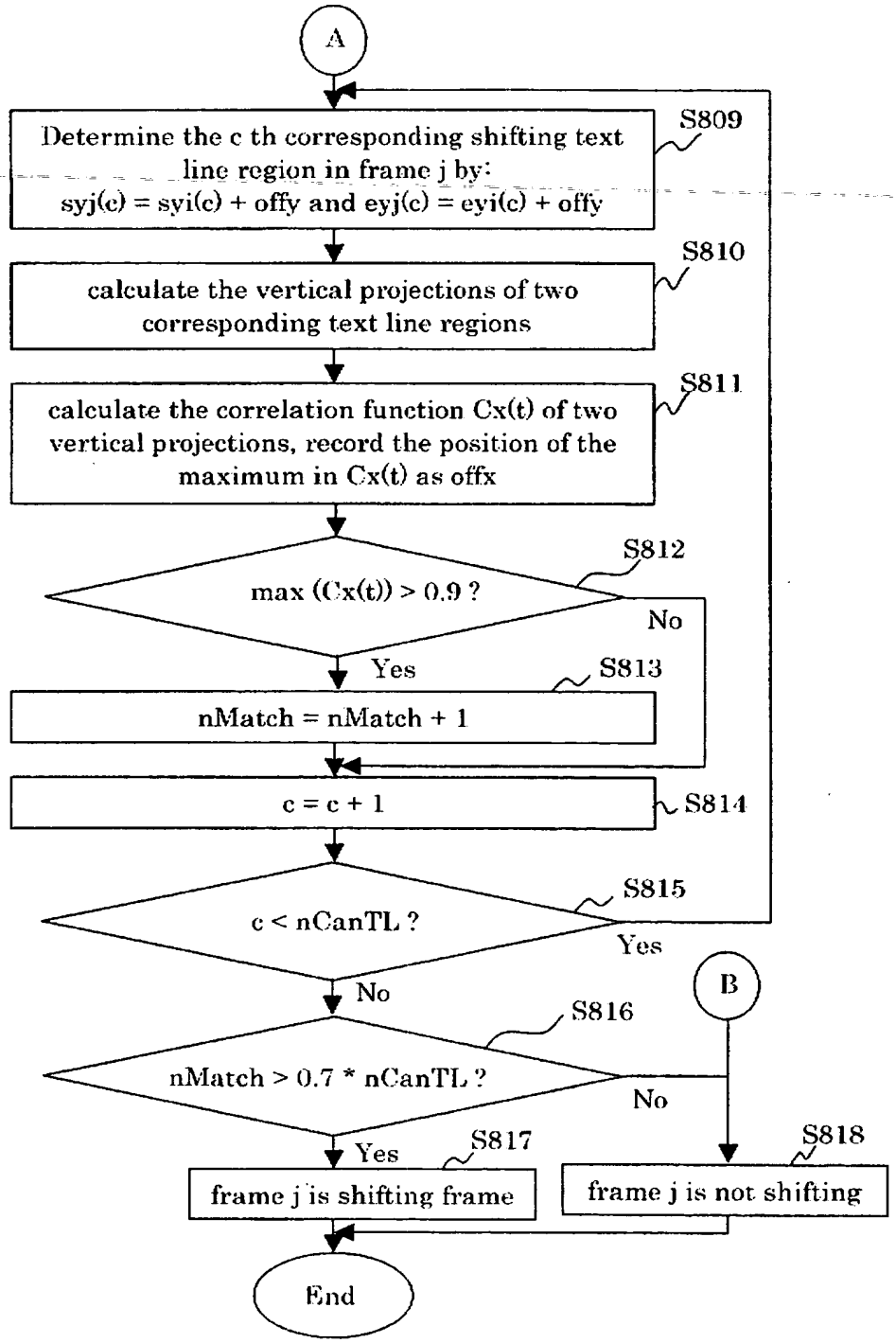
【図３２】

the flowchart of the operation
of the image shifting detection unit (No. 1)



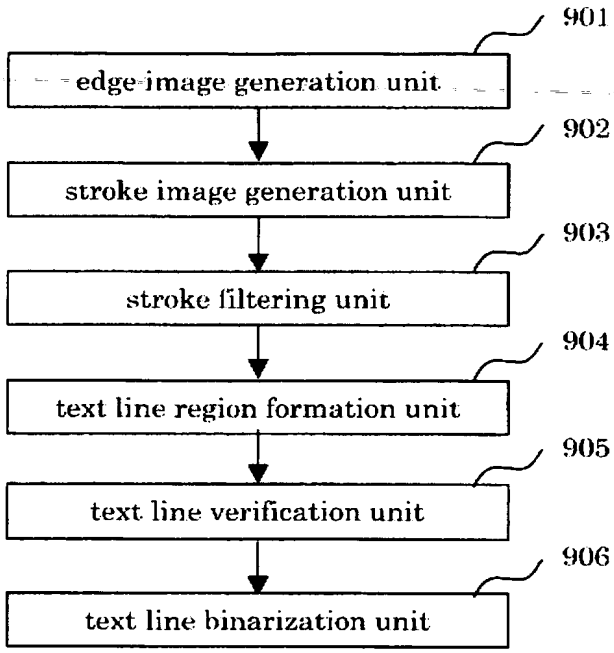
【図３３】

the flowchart of the operation
of the image shifting detection unit (No. 2)



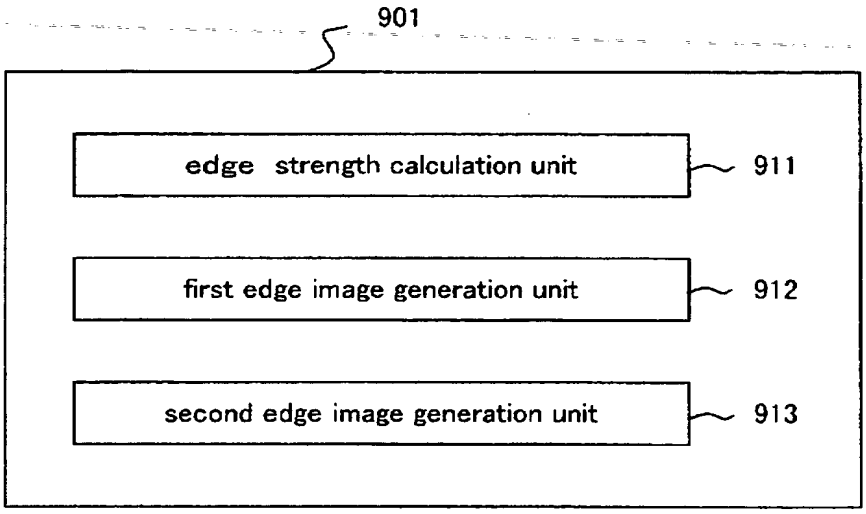
【図３４】

the drawing which shows the configuration of the text extraction apparatus according to the present invention



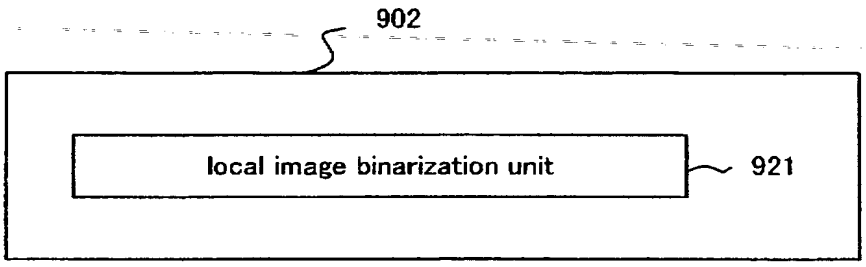
【図３５】

the drawing which shows the configuration of the edge image generation unit



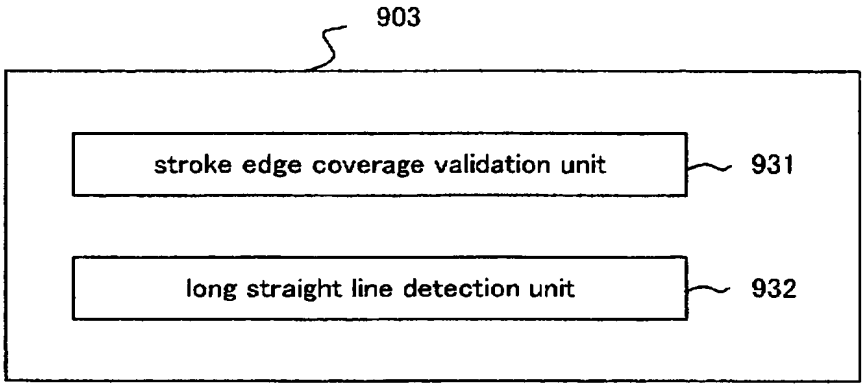
【図３６】

the drawing which shows
the configuration of the stroke
image generation unit



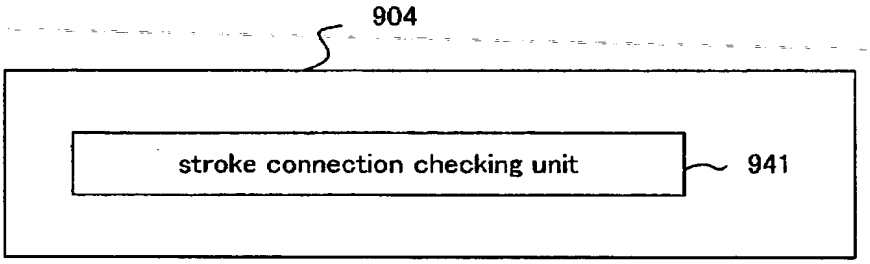
【図３７】

the drawing which shows
the configuration of the stroke filtering unit



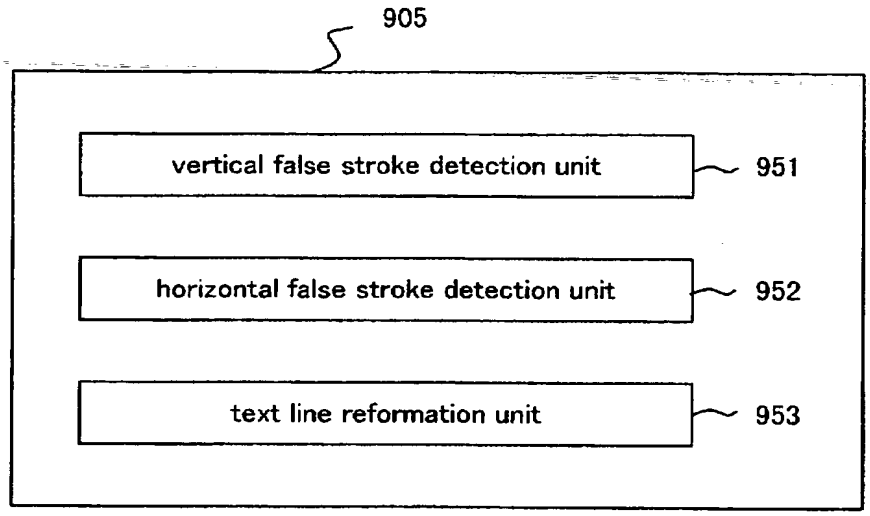
【図３８】

the drawing which shows
the configuration of the text
line region formation unit



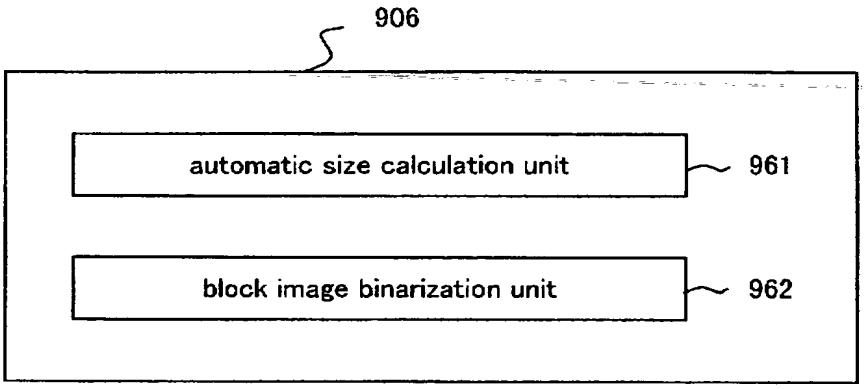
【図３９】

the drawing which shows
the configuration of the text
line verification unit



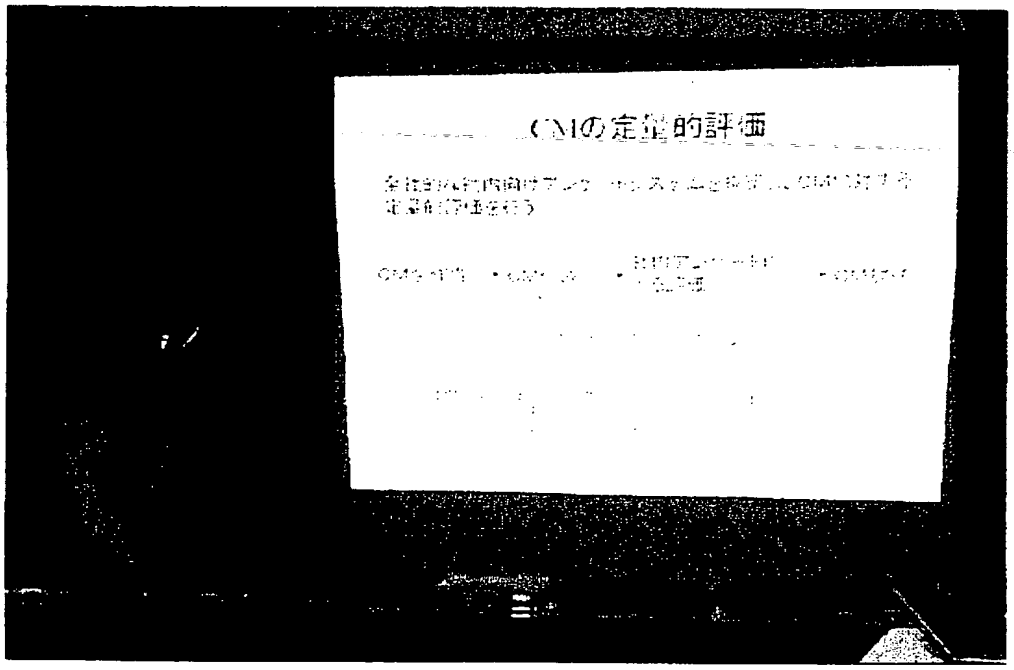
【図４０】

the drawing which shows the configuration of the text line binarization unit



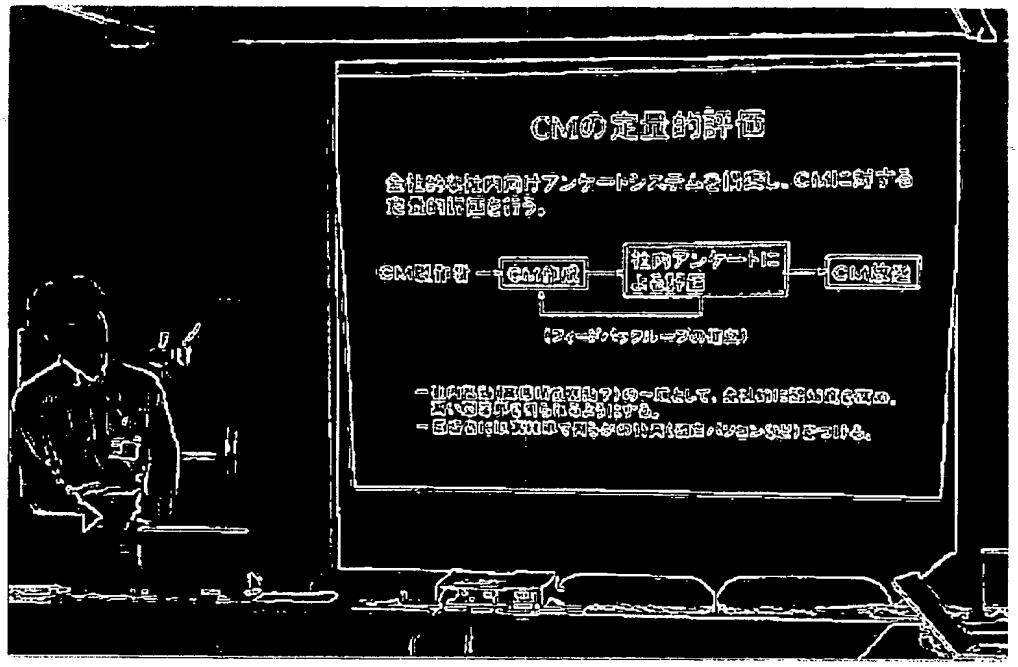
【図４１】

the drawing which shows
the original video frame for text extraction



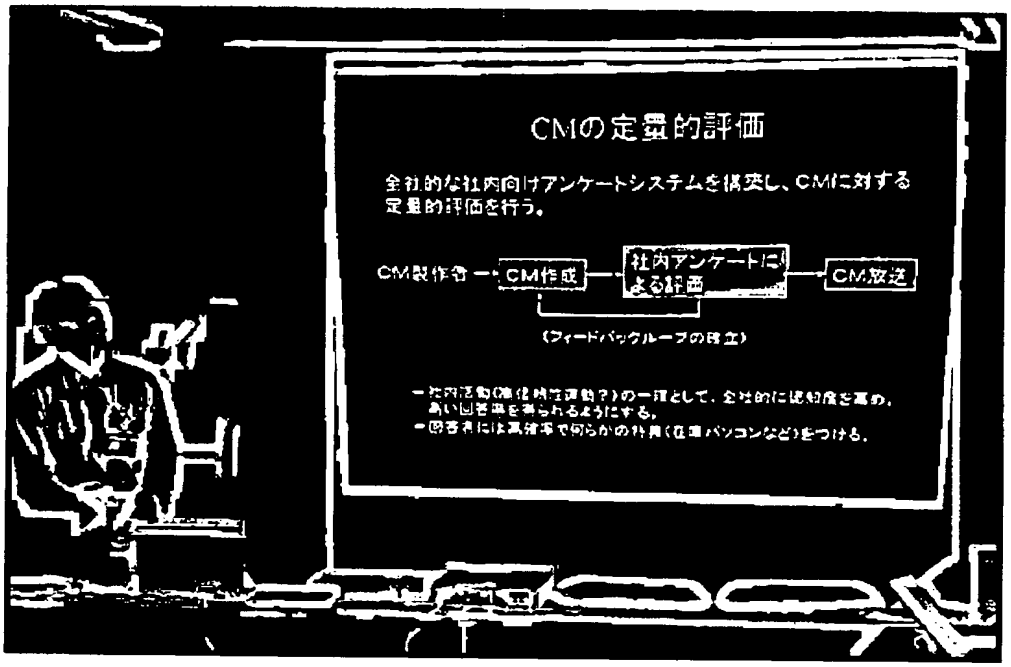
【図４２】

the drawing which shows
the result of edge image generation



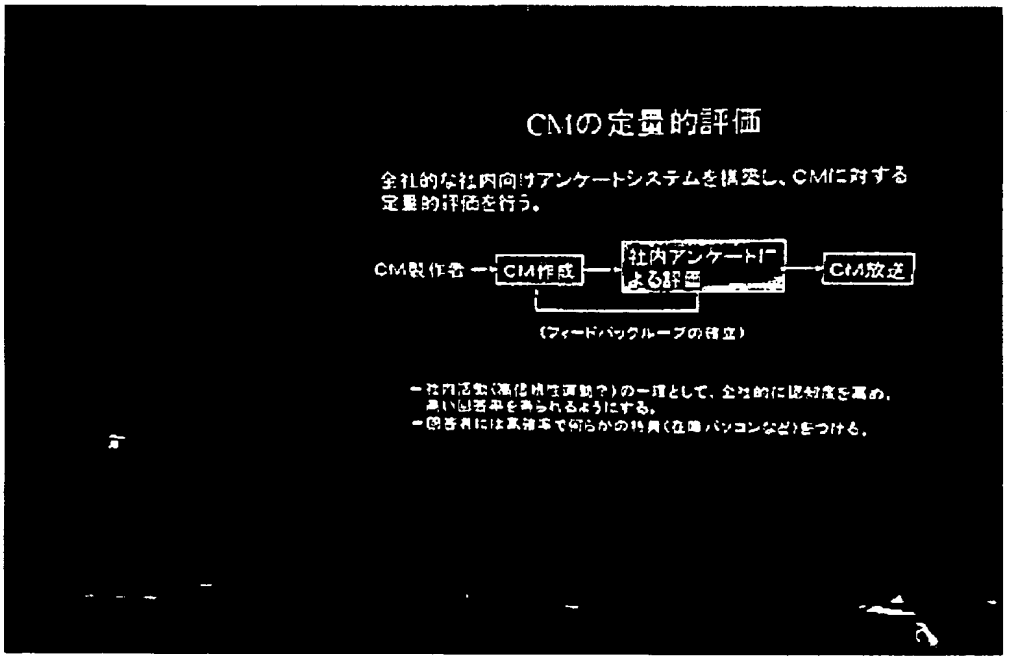
【図４３】

the drawing which shows
the result of stroke generation



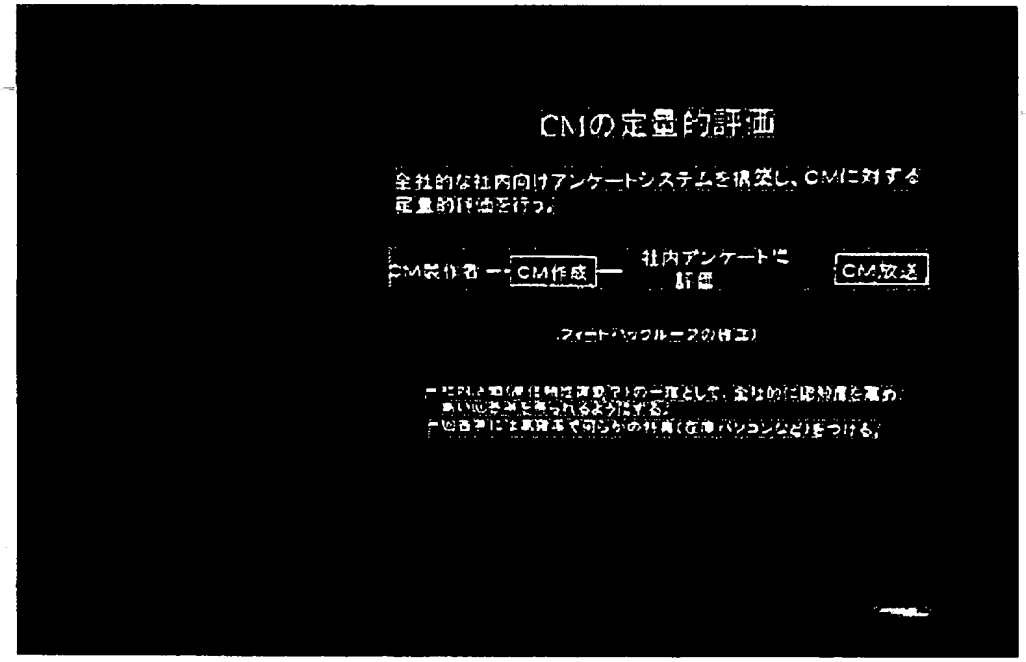
【図４４】

the drawing which shows the result of stroke filtering



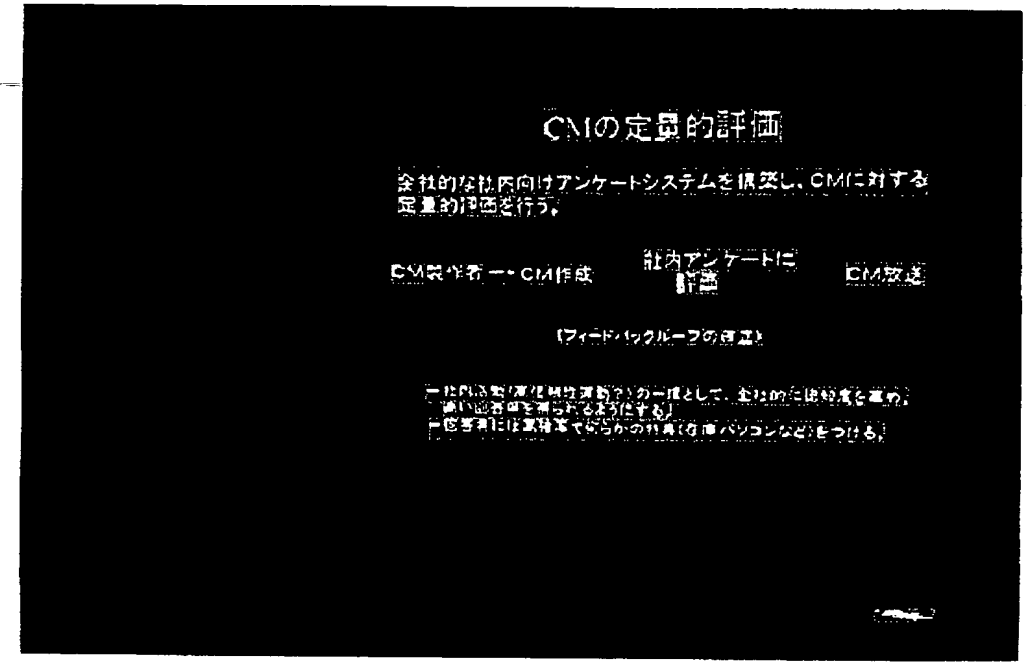
【図４５】

the drawing which shows
the result of text line region formation



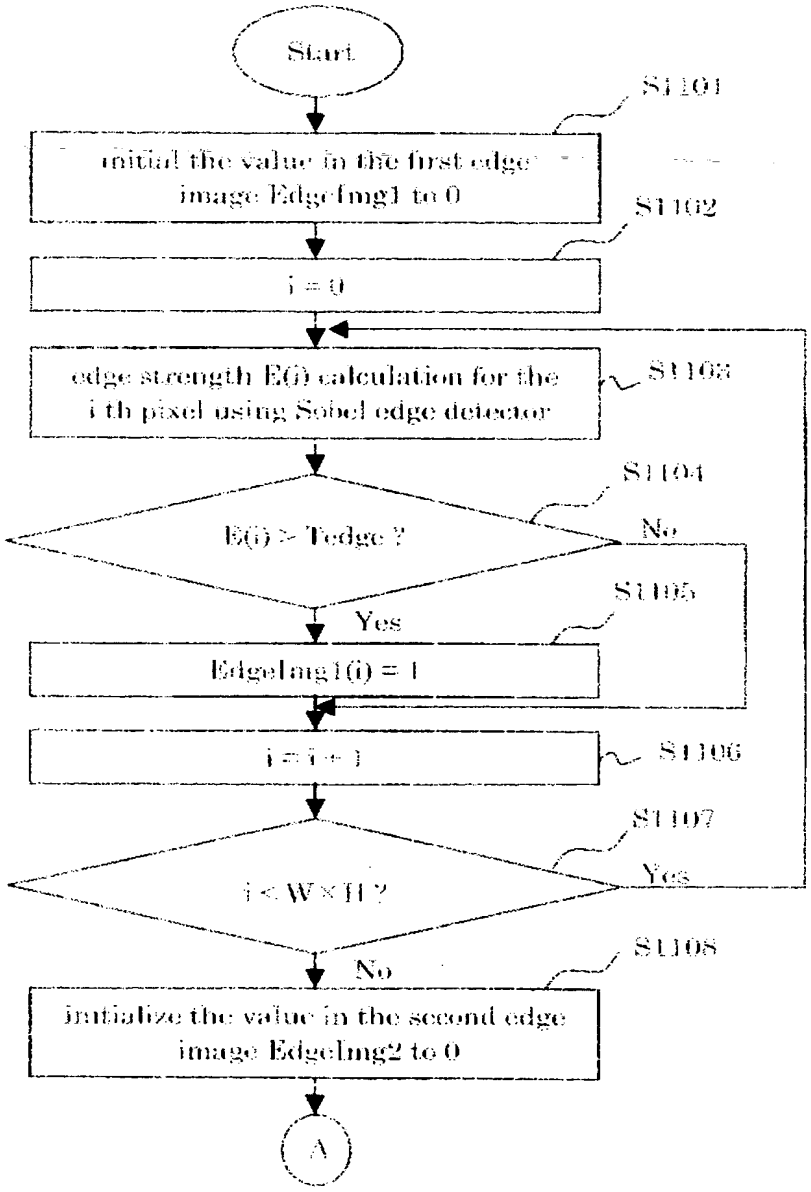
【図４６】

the drawing which shows
the final binarized text line regions



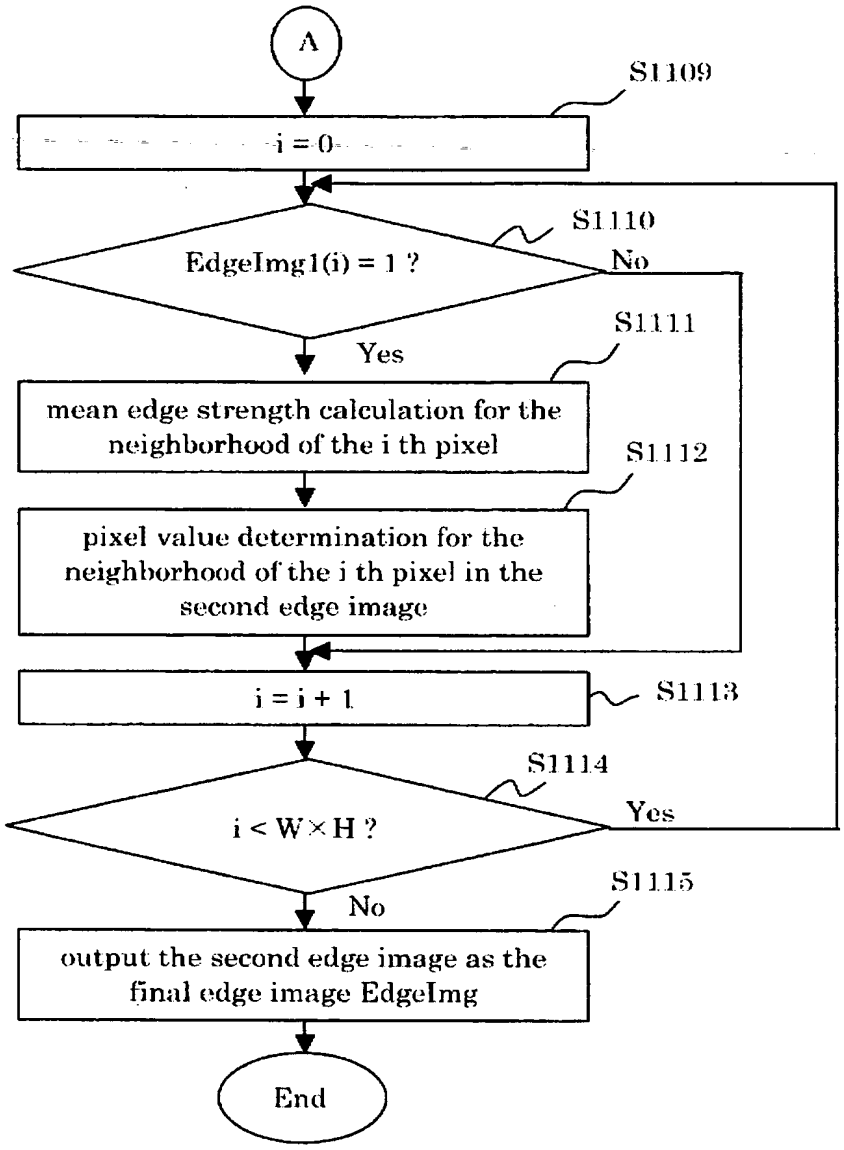
【図４７】

the flowchart of the operation
of the edge image generation unit (No. 1)



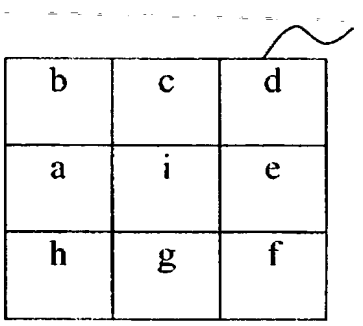
【図４８】

the flowchart of the operation
of the edge image generation unit (No. 2)



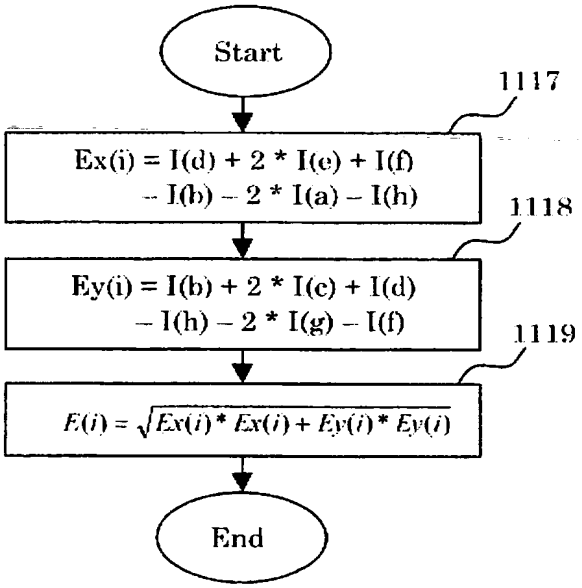
【図４９】

the drawing
which shows the arrangement
of the neighborhood of pixel i



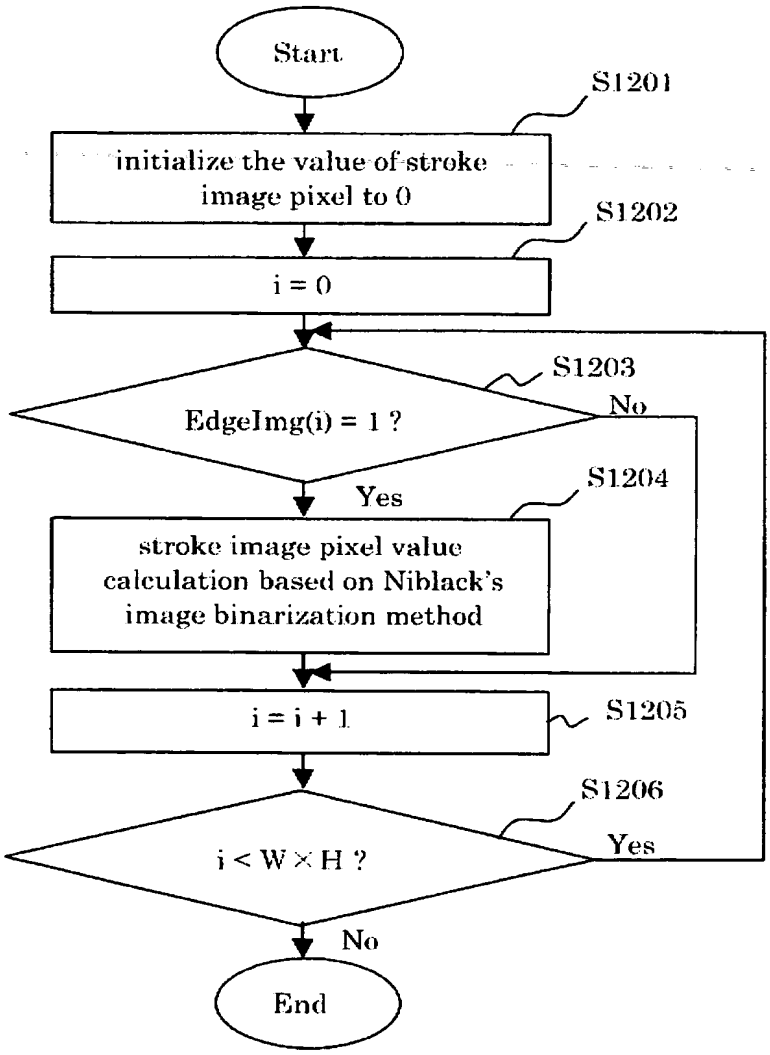
【図５０】

the flowchart of the operation
of the edge strength calculation unit



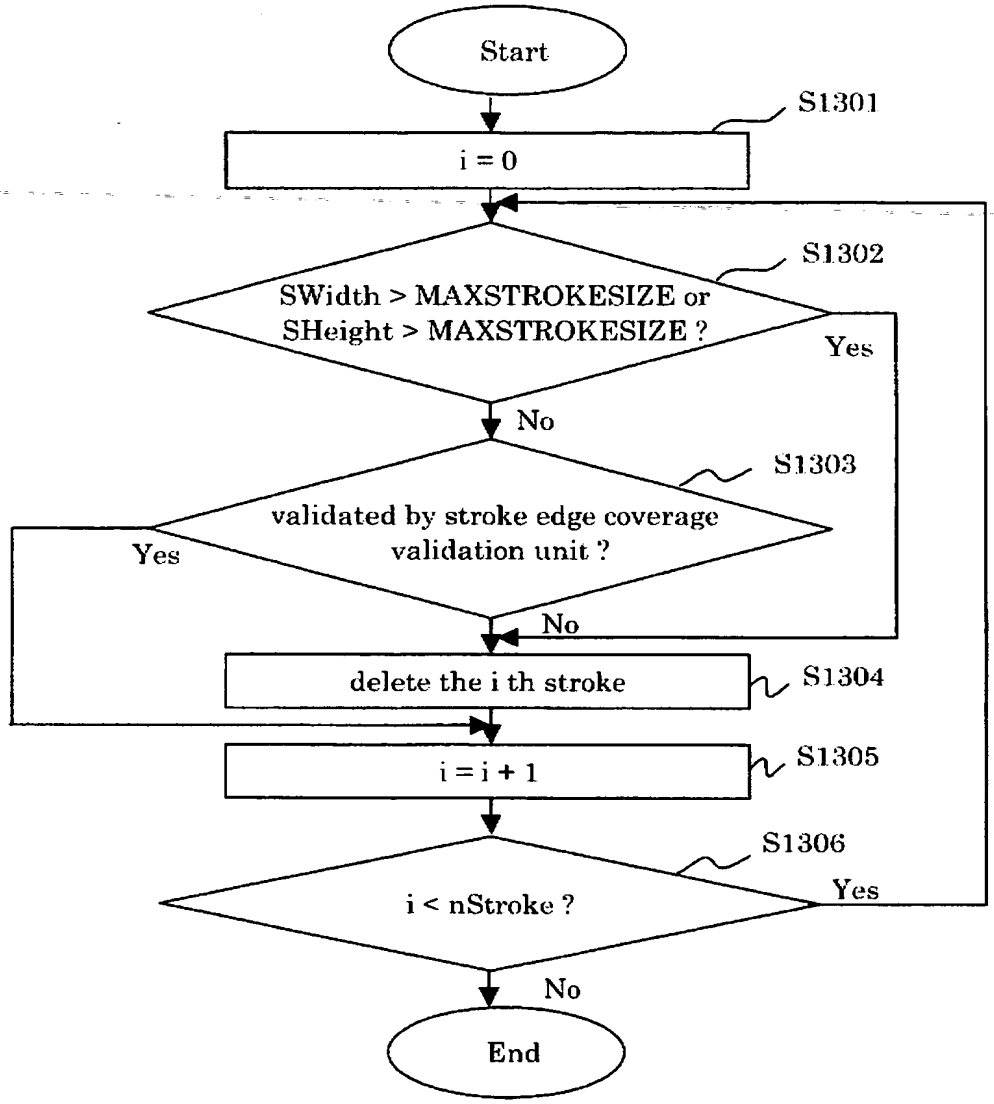
【図５１】

the flowchart of the operation
of the stroke image generation unit



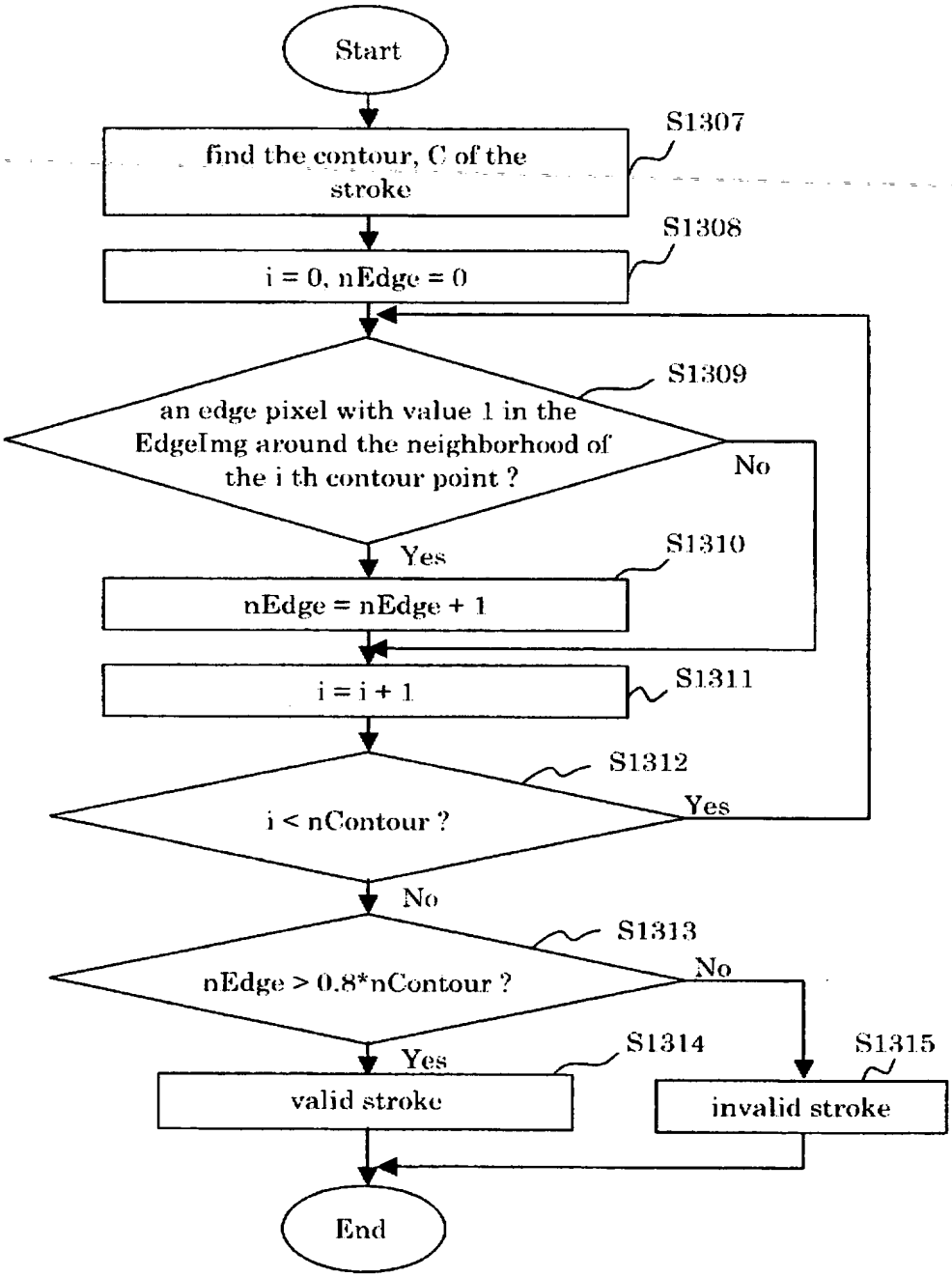
【図５２】

the flowchart of the operation
of the stroke filtering unit



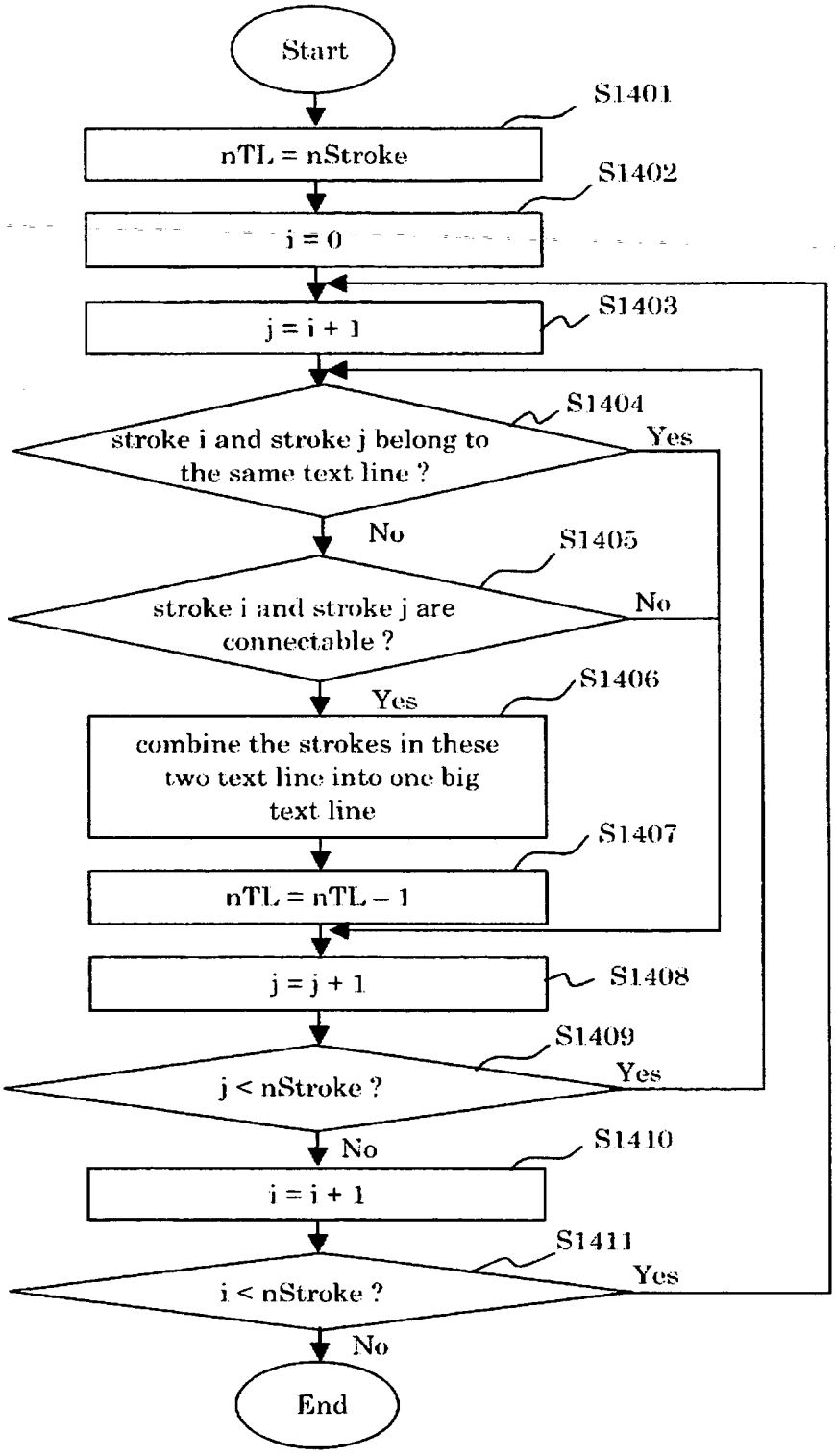
【図５３】

the flowchart of the operation
of the stroke edge coverage validation unit



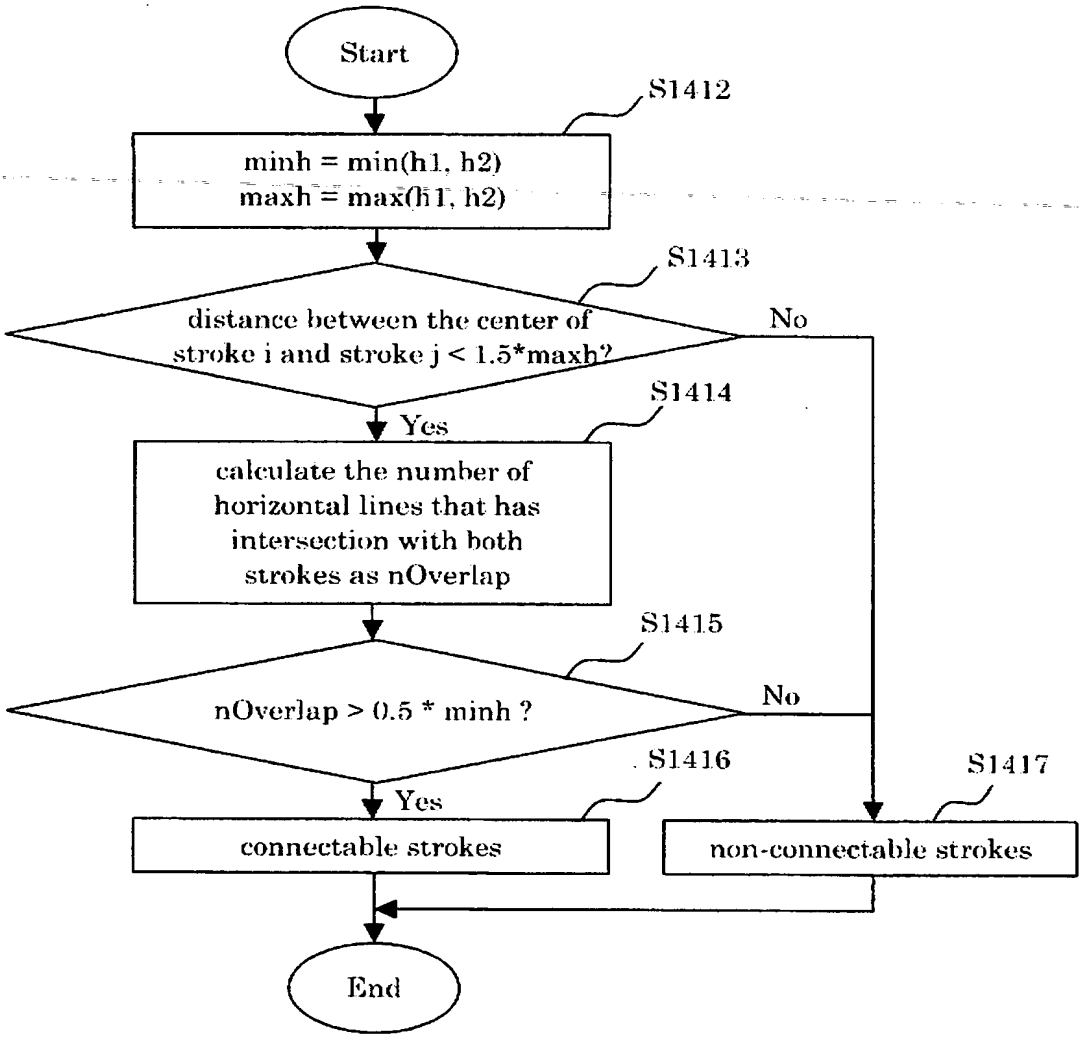
【図５４】

the flowchart of the operation
of the text line region formation unit



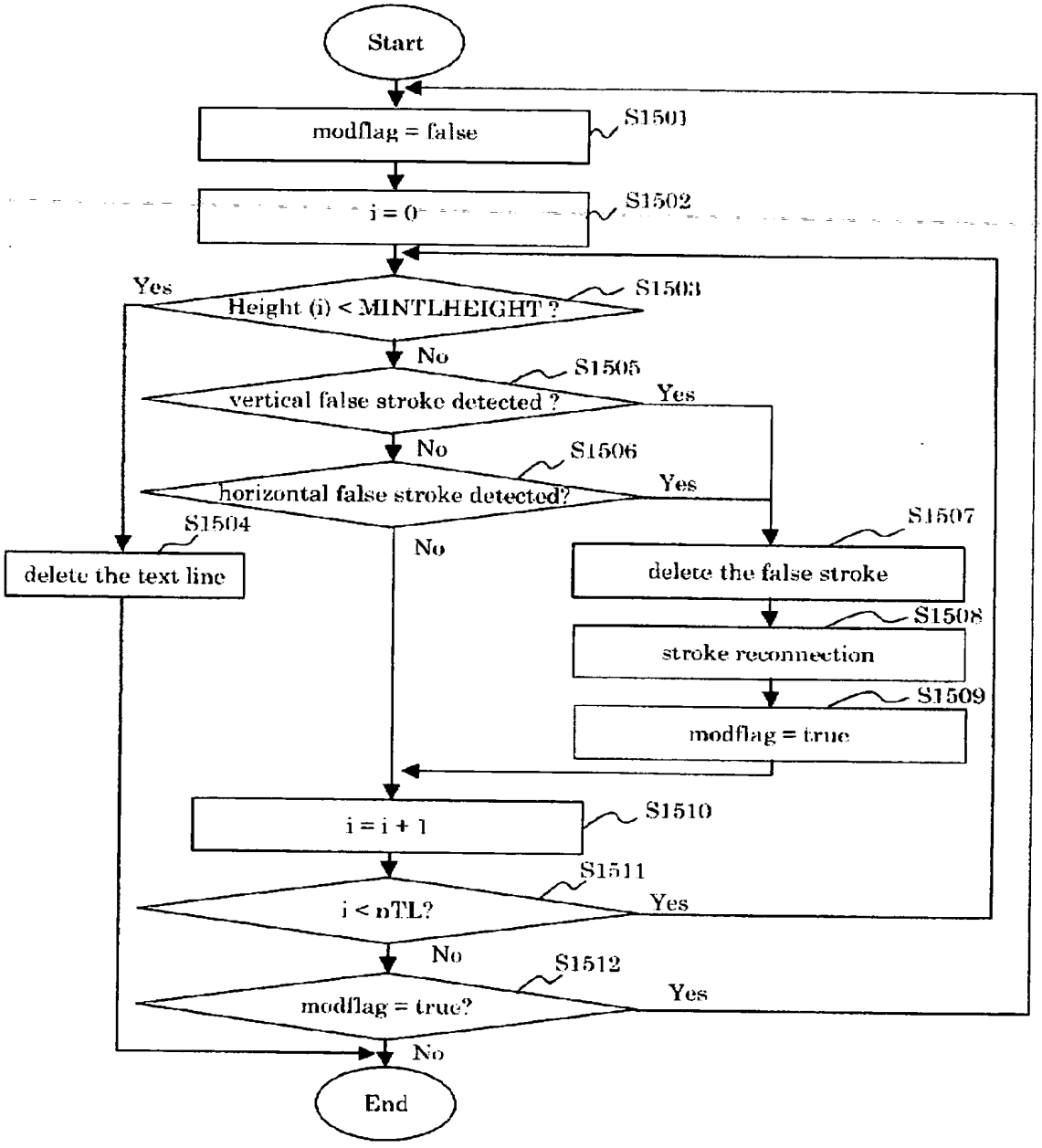
【図５５】

the flowchart of the operation
of the stroke connection checking unit



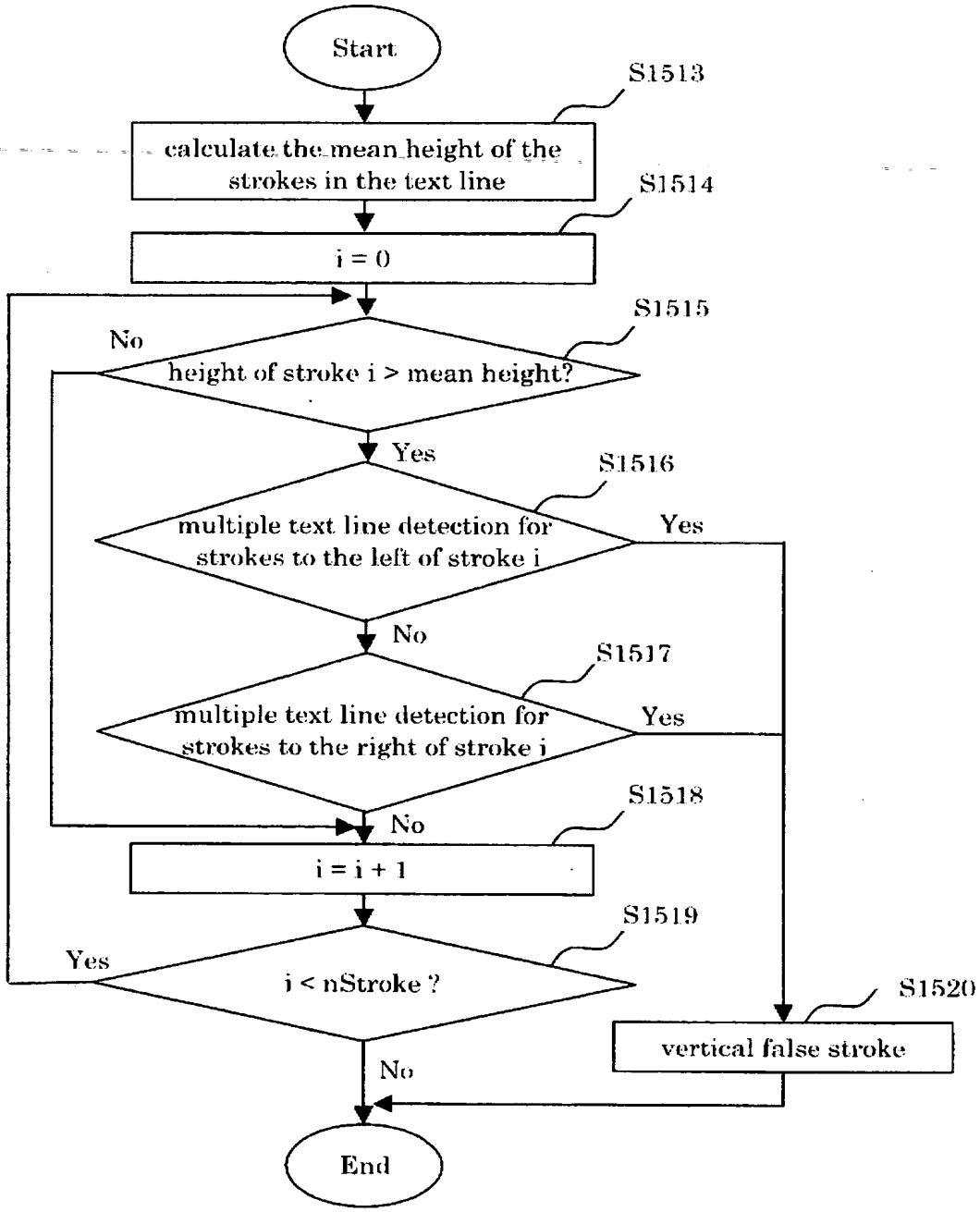
【図５６】

the flowchart of the operation
of the text line verification unit



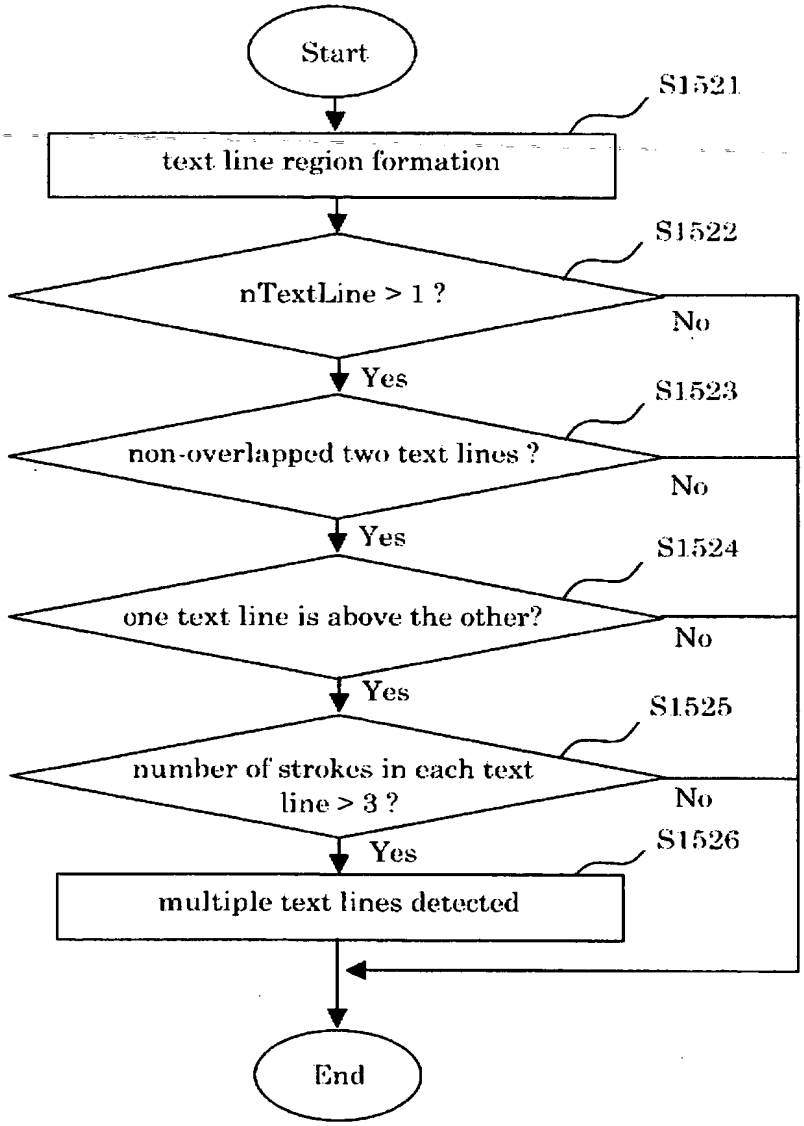
【図５７】

the flowchart of the operation
of the vertical false stroke detection unit



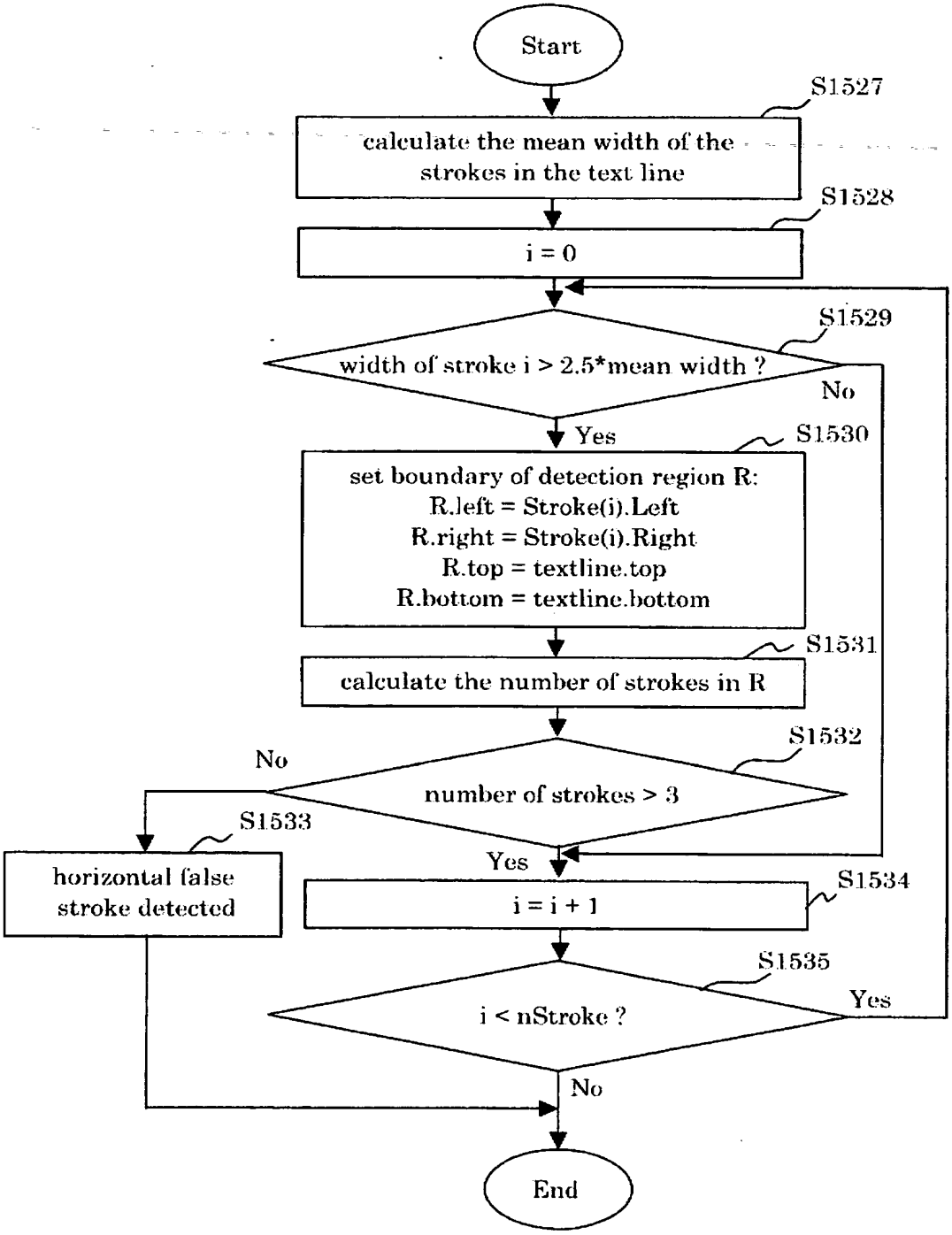
【図５８】

the flowchart
of multiple text line detection



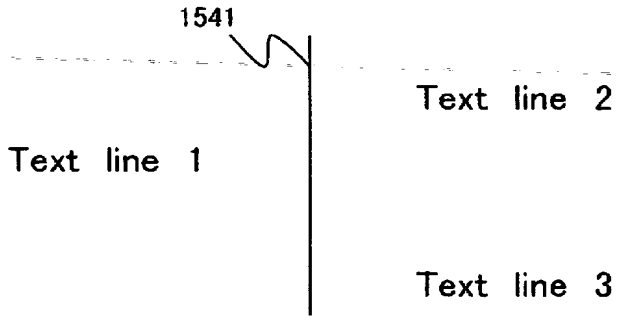
【図５９】

the flowchart of the operation
of the horizontal false stroke detection unit



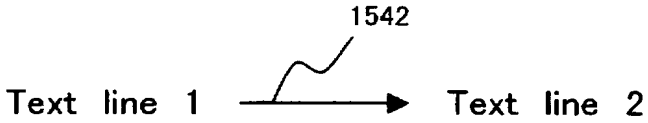
【図６０】

the drawing which
shows the first false stroke



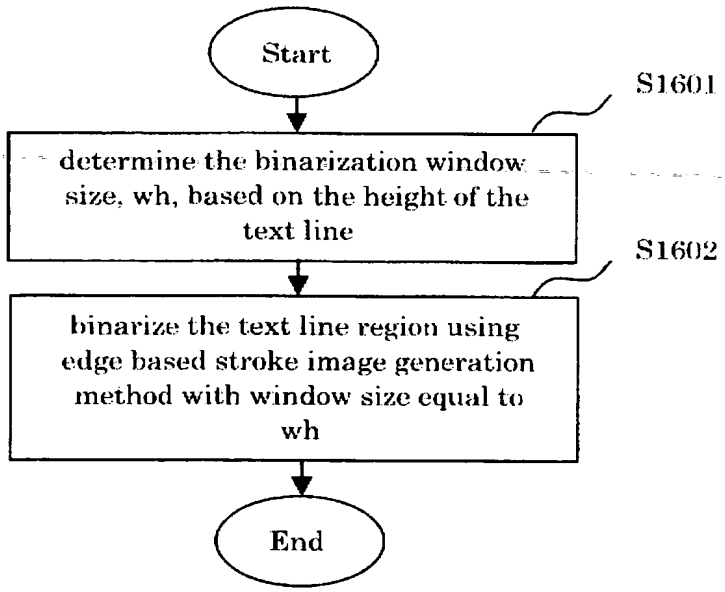
【図６１】

the drawing which
shows the second false stroke



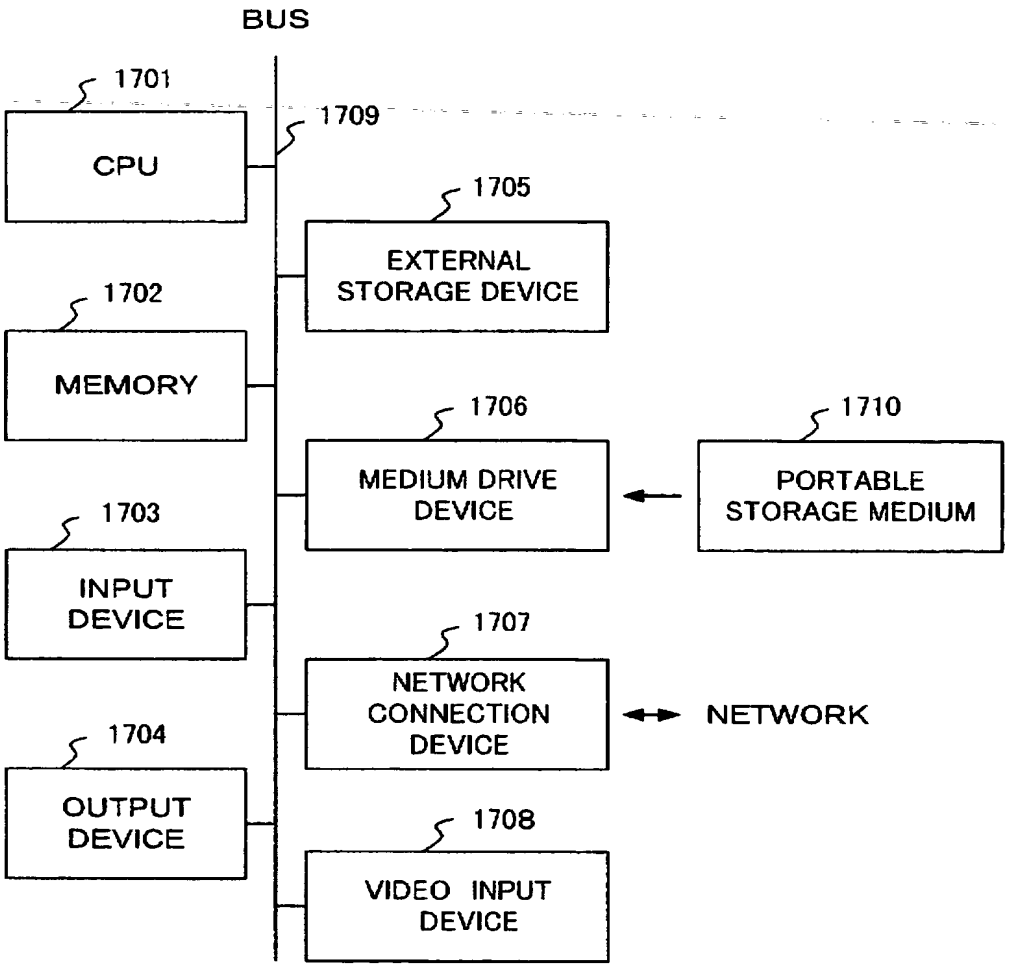
【図６２】

the flowchart of the operation of the text line binarization unit



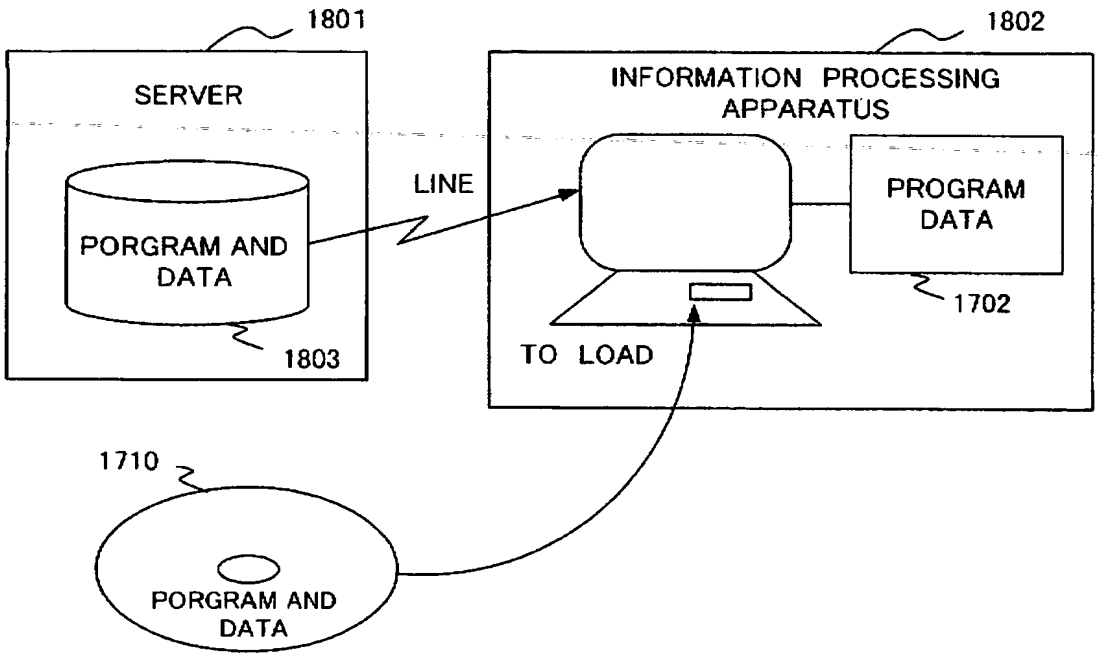
【図６３】

the drawing which shows
the configuration of an information
processing apparatus



【図６４】

the drawing which shows storage media



【書類名】　外国語要約書

【要約】

Problem to be solved

To select the candidate text change frames from a plurality of video frames in a fast speed and accurately detect the text region in the text change frame.

Solution

Video frames that contain text areas are selected from given video frames by removing redundant frames and non-text frames, the text areas in the selected frames are located by removing false strokes, and text lines in the text areas are extracted and binarized.

【選択図】　Fig.1

【書類名】　翻訳文提出書

【整理番号】　0252606

【提出日】　平成15年 2月24日

【あて先】　特許庁長官殿

【出願の表示】

【出願番号】　特願2002-378577

【特許出願人】

【識別番号】　000005223

【氏名又は名称】　富士通株式会社

【代理人】

【識別番号】　100074099

【弁理士】

【氏名又は名称】　大菅　義之

【電話番号】　03-3238-0031

【確認事項】　本書に添付した翻訳文は特願２００２－３７８５７７の正確な日本語への翻訳文であり、当該特許出願に記載されていない事項が本書に添付した翻訳文に記載されている場合には、当該出願が拒絶又は無効となる可能性があると承知していることを申し述べる。

【提出物件の目録】

【物件名】　外国語明細書の翻訳文　1

【物件名】　外国語図面の翻訳文　1

【物件名】　外国語要約書の翻訳文　1

【プルーフの要否】　要

【書類名】　　明細書

【発明の名称】　　ビデオテキスト処理装置

【特許請求の範囲】

【請求項１】　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するテキストチェンジフレーム検出装置であって、

前記与えられたビデオフレームから冗長なビデオフレームを除去する第１のフレーム除去手段と、

前記与えられたビデオフレームからテキスト領域を含まないビデオフレームを除去する第２のフレーム除去手段と、

前記与えられたビデオフレームから画像シフトに起因する冗長なビデオフレームを検出して除去する第３のフレーム除去手段と、

残されたビデオフレームを候補テキストチェンジフレームとして出力する出力手段と

を備えることを特徴とするテキストチェンジフレーム検出装置。

【請求項２】　与えられた画像から少なくとも１つのテキストライン領域を抽出するテキスト抽出装置であって、

前記与えられた画像のエッジ情報を生成するエッジ画像生成手段と、

前記エッジ情報を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成するストローク画像生成手段と、

前記エッジ情報を用いて前記二値画像から偽りのストロークを除去するストロークフィルタ手段と、

複数のストロークを１つのテキストライン領域に統合するテキストライン領域形成手段と、

前記テキストライン領域から偽りの文字ストロークを除去し、該テキストライン領域を改善するテキストライン検証手段と、

前記テキストライン領域の高さを用いて該テキストライン領域を二値化するテキストライン二値化手段と、

前記テキストライン領域の二値画像を出力する出力手段と

を備えることを特徴とするテキスト抽出装置。

【請求項３】　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムであって、

前記与えられたビデオフレームから冗長なビデオフレームを除去し、

前記与えられたビデオフレームからテキスト領域を含まないビデオフレームを除去し、

前記与えられたビデオフレームから画像シフトに起因する冗長なビデオフレームを検出して除去し、

残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【請求項４】　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムであって、

与えられたビデオフレームのうちの２つのビデオフレーム内の同じ位置にある２つの画像ブロックが、画像コンテンツの変化を示す能力のある有効ブロックペアであるか否かを決定し、

前記有効ブロックペアの２つの画像ブロックの類似度を計算して、該２つの画像ブロックが類似しているか否かを決定し、

有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて前記２つのビデオフレームが類似しているか否かを決定し、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【請求項５】　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムであって、

前記与えられたビデオフレームのうちの１つのビデオフレームの第１の二値画像を生成し、

前記第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決定し、

テキストライン領域毎に第２の二値画像を生成し、

前記第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の

総数に対する該テキストライン領域内のフォアグランド画素の数の充填率とを用いて、テキストライン領域の有効性を決定し、

１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、該１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームであるか否かを確認し、

前記非テキストフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【請求項６】　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムであって、

前記与えられたビデオフレームのうちの２つのビデオフレームの二値画像を生成し、

前記２つのビデオフレームの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定し、

前記横射影の相関を用いて、前記２つのビデオフレームの間における画像シフトの縦オフセットと、該２つのビデオフレームの縦方向の類似度とを決定し、

前記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、前記画像シフトの横オフセットと、該２つのビデオフレームの横方向の類似度とを決定し、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【請求項７】　与えられた画像から少なくとも１つのテキストライン領域を抽出するコンピュータのためのプログラムであって、

前記与えられた画像のエッジ情報を生成し、

前記エッジ情報を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ情報を用いて前記二値画像から偽りのストロークを除去し、

複数のストロークを１つのテキストライン領域に統合し、

前記テキストライン領域から偽りの文字ストロークを除去して、該テキストライン領域を改善し、

前記テキストライン領域の高さを用いて該テキストライン領域を二値化し、

前記テキストライン領域の二値画像を出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【請求項８】　与えられた画像から少なくとも１つのテキストライン領域を抽出するコンピュータのためのプログラムであって、

前記与えられた画像のエッジ画像を生成し、

前記エッジ画像を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ画像内のエッジを表す画素と前記候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、

前記重複率が所定のしきい値より大きければ該ストロークを有効ストロークと決定し、該重複率が該所定のしきい値より小さければ該ストロークを無効ストロークと決定し、

前記無効ストロークを除去し、

前記候補文字ストロークの二値画像内の残されたストロークの情報を出力する処理を前記コンピュータに実行させることを特徴とするプログラム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、ビデオ画像処理装置に関し、さらに詳しくはｅラーニングビデオのためのテキスト画像抽出装置に関する。テキストチェンジフレーム検出装置は、テキスト情報を含むビデオフレームを特定する。テキスト抽出装置は、そのビデオフレームからテキスト情報を抽出し、抽出したテキスト情報を認識のため光学式文字認識（ＯＣＲ）エンジンに渡す。

【０００２】

【従来の技術】

ビデオおよび画像内のテキスト検索は非常に重要な技術であり、格納領域削減

、ビデオおよび画像のインデックス付け、デジタルライブラリ等のように、多様な適用対象を有する。

【０００３】

本発明は、特定のタイプのビデオとして、大量のテキスト情報を含んでいるｅラーニングビデオに着目している。ビデオ内のテキストコンテンツを効率的に検索するためには、２つの技術が必要となる。ビデオ内でのテキストチェンジフレーム検出と画像からのテキスト抽出である。テキストチェンジフレームとは、ビデオ内でテキストコンテンツの変化を示すフレームである。第１の技術は、ビデオを高速に拾い読みしてテキスト領域を含むビデオフレームを選択する。そして、第２の技術は、そのビデオフレームからテキスト情報を抽出して、認識のためＯＣＲエンジンに渡す。

【０００４】

テキストチェンジフレーム検出技術は、場面チェンジフレーム検出技術の特別なケースであると考えられる。ビデオの内容の変化を示す場面チェンジフレームを検出する技術は、近年盛んに研究されている。いくつかの方法では、フレーム間の強度の違いに着目しており、いくつかの方法では、色ヒストグラムとテクスチャの違いに着目している。しかしながら、これらの方法は、特にｅラーニングの分野においては、ビデオ内でのテキストチェンジフレーム検出に適していない。

【０００５】

典型的なｅラーニングビデオの例として、ビデオフレームがしばしばスライド画像を含んでいるプレゼンテーションビデオを採り上げてみる。スライド画像の例としては、PowerPoint（登録商標）画像やプロジェクタからのフィルム画像がある。スライドの内容の変化は、色およびテクスチャの劇的な変化を引き起こすことにはならない。また、ビデオカメラの焦点は、話中にしばしばスライド画像内であちこち移動して、画像シフトを引き起こす。話し手が自分のスライドを動かすときにも、画像シフトが発生する。従来の方法では、これらの内容シフトフレームが場面チェンジフレームとしてマークされることになる。従来の方法のもう１つの欠点は、あるフレームがテキスト情報を含んでいるか否かを直接判定で

きないことである。

【０００６】

ビデオからテキストチェンジフレームを抽出するもう１つの方法は、ビデオ内のフレーム毎にテキスト抽出法を実行して、内容が変化したか否かを判定することである。この戦略の問題点は、非常に時間がかかるということである。

【０００７】

テキストチェンジフレームが検出された後、テキスト抽出法が用いて、そのフレームからテキストラインが抽出される。ビデオおよび静止画像からテキストラインを抽出するために多くの方法が提唱されている（例えば、非特許文献１および２参照）。

【０００８】

また、本分野に関連するいくつかの特許も公開されている（例えば、特許文献１、２、および３参照）。

これらの方法によれば、ｅラーニングにおけるビデオフレームを取り扱うときに問題が生じることになる。ｅラーニングビデオ画像内の文字は常に極小サイズであり、文字の境界も非常に不明瞭であり、テキスト領域の周辺には、テキストラインのバウンディングボックス、人間の体による陰影や隠蔽等のように、多くの障害がある。

【０００９】

【非特許文献１】

V. Wu, R. Manmatha, and E. M. Riseman, “TextFinder: An Automatic System to Detect and Recognize Text in Images,” IEEE transactions on Pattern Analysis and Machine Intelligence, VOL. 21, NO. 11, pp. 1224-1229, November, 1999.

【非特許文献２】

T. Sato, T. Kanade, E. Hughes, M. Smith, and S. Satoh, “Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Captions,” ACM Multimedia Systems Special Issue on Video Libraries, February, 1998.

【特許文献１】

米国特許第６，３６６，６９９号明細書

【特許文献２】

米国特許第５，４６５，３０４号明細書

【特許文献３】

米国特許第５，３０７，４２２号明細書

【００１０】

【発明が解決しようとする課題】

しかしながら、上述した従来のビデオ画像処理には、次のような問題がある。

ビデオ内のフレーム毎にテキスト抽出法を実行して、内容が変化したか否かを判定するのには、非常に時間がかかる。

【００１１】

ｅラーニングビデオ画像内の文字は常に極小サイズであり、文字の境界も非常に不明瞭であり、テキスト領域の周辺には多くの障害がある。したがって、従来のテキスト抽出法では、最終二値画像内に多くの偽りの文字ストロークが残ってしまい、後続のＯＣＲ段階で誤った認識結果を生じさせることになる。

【００１２】

本発明の課題は、正しいテキストチェンジフレームの総数に対する抽出された正しいテキストチェンジフレームの数の比率として定義されるリコール率を高く維持したままで、複数のビデオフレームから候補テキストチェンジフレームを高速に選択することである。

【００１３】

本発明のもう１つの課題は、テキストチェンジフレーム内のテキスト領域を効率的に検出し、偽りの文字ストロークをできるだけ多く除去して、テキストライン毎の二値画像を提供するスキームを提供することである。

【００１４】

【課題を解決するための手段】

上記課題は、ビデオ内のテキストフレームを高速に選択するテキストチェンジフレーム検出装置と、テキストフレーム内のテキストラインを抽出するテキスト

抽出装置とを備え、ビデオ内の全フレームからテキストコンテンツを含むフレームを高速に選択し、テキストフレーム内の各テキストラインの領域をマークして、テキストラインをバイナリ形式で出力するビデオテキスト処理装置により達成される。

【００１５】

第１のテキストチェンジフレーム検出装置は、第１のフレーム除去手段、第２のフレーム除去手段、第３のフレーム除去手段、および出力手段を備え、与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択する。第１のフレーム除去手段は、与えられたビデオフレームから冗長なビデオフレームを除去する。第２のフレーム除去手段は、与えられたビデオフレームからテキスト領域を含まないビデオフレームを除去する。第３のフレーム除去手段は、与えられたビデオフレームから画像シフトに起因する冗長なビデオフレームを検出して除去する。出力手段は、残されたビデオフレームを候補テキストチェンジフレームとして出力する。

【００１６】

第２のテキストチェンジフレーム検出装置は、画像ブロック有効化手段、画像ブロック類似度計測手段、フレーム類似度判定手段、および出力手段を備え、与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択する。

【００１７】

画像ブロック有効化手段は、与えられたビデオフレームのうちの２つのビデオフレーム内の同じ位置にある２つの画像ブロックが、画像コンテンツの変化を示す能力のある有効ブロックペアであるか否かを決定する。画像ブロック類似度計測手段は、有効ブロックペアの２つの画像ブロックの類似度を計算して、それらの２つの画像ブロックが類似しているか否かを決定する。フレーム類似度判定手段は、有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて２つのビデオフレームが類似しているか否かを決定する。出力手段は、類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する。

【００１８】

第３のテキストチェンジフレーム検出装置は、高速簡易画像二値化手段、テキストライン領域決定手段、再二値化手段、テキストライン確認手段、テキストフレーム検証手段、および出力手段を備え、与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択する。高速簡易画像二値化手段は、与えられたビデオフレームのうちの１つのビデオフレームの第１の二値画像を生成する。テキストライン領域決定手段は、第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決定する。再二値化手段は、テキストライン領域毎に第２の二値画像を生成する。テキストライン確認手段は、第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の総数に対するそのテキストライン領域内のフォアグランド画素の数の充填率とを用いて、テキストライン領域の有効性を決定する。テキストフレーム検証手段は、１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームであるか否かを確認する。出力手段は、非テキストフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する。

【００１９】

第４のテキストチェンジフレーム検出装置は、高速簡易画像二値化手段、テキストライン縦位置決定手段、縦シフト検出手段、横シフト検出手段、および出力手段を備え、与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択する。高速簡易画像二値化手段は、与えられたビデオフレームのうちの２つのビデオフレームの二値画像を生成する。テキストライン縦位置決定手段は、上記２つのビデオフレームの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定する。縦シフト検出手段は、横射影の相関を用いて、上記２つのビデオフレームの間における画像シフトの縦オフセットと、それらの２つのビデオフレームの縦方向の類似度とを決定する。横シフト検出手段は、上記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、画像シフトの横オフセットと、それらの２つのビデオフレームの横方向の類似度とを決定する。出力手段は、類似ビデオフレームを除去した後に残さ

れたビデオフレームを候補テキストチェンジフレームとして出力する。

【００２０】

テキストチェンジフレーム検出装置によりビデオ内の候補テキストチェンジフレームが検出された後、フレーム毎の画像がテキスト抽出のためにテキスト抽出装置に送られる。

【００２１】

第１のテキスト抽出装置は、エッジ画像生成手段、ストローク画像生成手段、ストロークフィルタ手段、テキストライン領域形成手段、テキストライン検証手段、テキストライン二値化手段、および出力手段を備え、与えられた画像から少なくとも１つのテキストライン領域を抽出する。

【００２２】

エッジ画像生成手段は、与えられた画像のエッジ情報を生成する。ストローク画像生成手段は、エッジ情報を用いて与えられた画像内の候補文字ストロークの二値画像を生成する。ストロークフィルタ手段は、エッジ情報を用いて上記二値画像から偽りのストロークを除去する。テキストライン領域形成手段は、複数のストロークを１つのテキストライン領域に統合する。テキストライン検証手段は、テキストライン領域から偽りの文字ストロークを除去し、そのテキストライン領域を改善する。テキストライン二値化手段は、上記テキストライン領域の高さを用いてそのテキストライン領域を二値化する。出力手段は、上記テキストライン領域の二値画像を出力する。

【００２３】

第２のテキスト抽出装置は、エッジ画像生成手段、ストローク画像生成手段、ストロークフィルタ手段、および出力手段を備え、与えられた画像から少なくとも１つのテキストライン領域を抽出する。エッジ画像生成手段は、与えられた画像のエッジ画像を生成する。ストローク画像生成手段は、エッジ画像を用いて与えられた画像内の候補文字ストロークの二値画像を生成する。ストロークフィルタ手段は、エッジ画像内のエッジを表す画素と上記候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、重複率が所定のしきい値より大きければそのストロークを有効ストロークと決定し、重複率が所定のしきい値

より小さければそのストロークを無効ストロークと決定して、無効ストロークを除去する。出力手段は、上記候補文字ストロークの二値画像内の残されたストロークの情報を出力する。

【００２４】

テキスト抽出装置によりテキストライン領域が抽出された後、それらは認識のためにＯＣＲエンジンに送られる。

【００２５】

【発明の実施の形態】

以下、図面を参照しながら、本発明の実施の形態を詳細に説明する。

図１は、本発明のビデオテキスト処理装置の構成を示している。装置の入力は、既存のビデオデータ１０１またはテレビ（ＴＶ）ビデオカメラ１０２からの生のビデオストリームであり、入力されたビデオデータは、まずビデオ分解部１０３により連続したフレームに分解される。そして、テキストチェンジフレーム検出装置１０４を用いて、ビデオフレーム内の候補テキストチェンジフレームが見つけられる。テキストチェンジフレーム検出装置は、総処理時間を大幅に削減することになる。その後、テキスト抽出装置１０５に、候補テキストチェンジフレーム毎にフレーム内のテキストライン（テキスト領域）を検出させ、そのテキストラインの画像をさらなるＯＣＲ処理のためにデータベース１０６に出力させる。

【００２６】

図２は、図１のビデオテキスト処理装置の処理フローチャートを示している。Ｓ２０１の処理はビデオ分解部１０３により行われ、Ｓ２０２からＳ２０４までの処理はテキストチェンジフレーム検出装置１０４により行われ、Ｓ２０５からＳ２１０までの処理はテキスト抽出装置１０５により行われる。

【００２７】

まず、入力されたビデオが連続したフレームに分解される（Ｓ２０１）。そして、フレーム類似度計測が行われ、２つの隣同士のフレームの類似度が計測される（Ｓ２０２）。それらの２つのフレームが類似していれば、２番目のフレームが除去される。次に、テキストフレーム検出および検証が行われ、Ｓ２０２の処

理で残されたフレームがテキストラインを含むか否かが判定される（Ｓ２０３）。フレームがテキストラインを含んでいなければ、そのフレームは除去される。さらに画像シフト検出が行われ、２つのフレーム内に画像シフトが存在するか否かが決定される（Ｓ２０４）。もしそうであれば、２番目のフレームが除去される。テキストチェンジフレーム検出装置１０４の出力は、一群の候補テキストチェンジフレームとなる。

【００２８】

候補テキストチェンジフレーム毎にエッジ画像生成が行われ、フレームのエッジ画像が生成される（Ｓ２０５）。そして、ストローク生成が行われ、エッジ情報に基づいてストローク画像が生成される（Ｓ２０６）。次に、ストロークフィルタリングが行われ、エッジ情報に基づいて偽りのストロークが除去される（Ｓ２０７）。さらにテキストライン領域形成が行われ、個々のストロークが接続されて１つのテキストラインになる（Ｓ２０８）。その後、テキストライン検証が行われ、テキストライン内の偽りのストロークが除去されて、テキストラインが改善される（Ｓ２０９）。最後に、テキストライン二値化が行われ、最終的な二値画像が生成される（Ｓ２１０）。最終出力は、認識のためにＯＣＲエンジンにより処理されることになる一連の二値テキストライン画像である。

【００２９】

図３は、図１のテキストチェンジフレーム検出装置１０４の構成を示している。入力されたビデオフレームは、まず重複しているフレームを削除するためにフレーム類似度計測部３０１に送られ、テキストフレーム検出検証部３０２を用いて、テキスト情報を含むフレームか否かがチェックされる。次に、画像シフト検出部３０３を用いて、画像シフトに起因する冗長なフレームが除去される。フレーム類似度計測部３０１、テキストフレーム検出検証部３０２、および画像シフト検出部３０３は、それぞれ第１、第２、および第３のフレーム除去手段に対応する。テキストチェンジフレーム検出装置１０４は、ｅラーニングビデオ内のテキストチェンジフレームを検出するのに非常に適している。この装置は、テキスト領域を含んでいないビデオフレームとともに、重複しているビデオフレームおよびシフトしているビデオフレームを、高いリコール率を維持しながら非常に高

速に除去することができる。

【００３０】

図４は、図３のフレーム類似度計測部３０１の構成を示している。フレーム類似度計測部３０１は、画像ブロック有効化部３１１、画像ブロック類似度計測部３１２、およびフレーム類似度判定部３１３を含む。画像ブロック有効化部３１１は、２つのビデオフレーム内の同じ位置にある２つの画像ブロックが有効ブロックペアであるか否かを決定する。有効ブロックペアとは、画像内容の変化を示す能力のある画像ブロックペアである。画像ブロック類似度計測部３１２は、有効ブロックペアの２つの画像ブロックの類似度を計算し、それらの２つの画像ブロックが類似しているか否かを決定する。フレーム類似度判定部３１３は、有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて、２つのビデオフレームが類似しているか否かを決定する。フレーム類似度計測部３０１によれば、ビデオフレームから重複しているフレームが効率的に検出され、除去される。

【００３１】

図５は、図３のテキストフレーム検出検証部３０２の構成を示している。テキストフレーム検出検証部３０２は、高速簡易画像二値化部３２１、テキストライン領域決定部３２２、再二値化部３２３、テキストライン確認部３２４、およびテキストフレーム検証部３２５を含む。高速簡易画像二値化部３２１は、ビデオフレームの第１の二値画像を生成する。テキストライン領域決定部３２２は、第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決定する。再二値化部３２３は、テキストライン領域毎に第２の二値画像を生成する。テキストライン確認部３２４は、第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の総数に対するそのテキストライン領域内のフォアグランド画素の数の充填率とを用いて、テキストライン領域の有効性を決定する。テキストフレーム検証部３２５は、１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームであるか否かを確認する。テキストフレーム検出検証部３０２によれば、ビデオフレームから非テキストフレームが高速に検出され、除去

される。

【００３２】

図６は、図３の画像シフト検出部３０３の構成を示している。画像シフト検出部３０３は、高速簡易画像二値化部３３１、テキストライン縦位置決定部３３２、縦シフト検出部３３３、および横シフト検出部３３４を含む。高速簡易画像二値化部３３１は、２つのビデオフレームの二値画像を生成する。テキストライン縦位置決定部３３２は、それらの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定する。縦シフト検出部３３３は、横射影の相関を用いて、上記２つのビデオフレームの間における画像シフトの縦オフセットと、それらの２つのビデオフレームの縦方向の類似度とを決定する。横シフト検出部３３４は、上記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、画像シフトの横オフセットと、それらの２つのビデオフレームの横方向の類似度とを決定する。画像シフト検出部３０３によれば、ビデオフレームから画像シフトに起因する冗長なフレームが高速に検出され、除去される。

【００３３】

図７および８は、同じテキストコンテンツを有する２つのフレームを示している。図９は、これらの２つのフレームに対するフレーム類似度計測部３０１の処理結果を示している。図９の白のボックスは、有効ブロックペアに含まれるブロックであり、かつ、内容の変化を示す能力のある、すべての有効画像ブロックを示している。実線のボックスは類似画像ブロックを表し、破線のボックスは非類似画像ブロックを表す。有効ブロックの数に対する類似画像ブロックの数の比は、所定のしきい値より大きいので、これらの２つの画像は類似しているとみなされ、２番目のフレームが除去される。

【００３４】

図１０は、図４のフレーム類似度計測部３０１の動作フローチャートである。０番目の瞬間の０番目のフレームから比較が開始され（Ｓ５０１）、現在のｉ番目のフレームが、ＳＴＥＰ個のフレームのフレーム間隔を有するｊ番目のフレームと比較される（Ｓ５０２）。２つのフレームの比較においてｉ番目のフレームがｊ番目のフレームと類似していれば（Ｓ５０３）、現在のフレーム位置がｊ番

目のフレームにジャンプして、比較のためのＳ５０２およびＳ５０３の処理が繰り返される。

【００３５】

２つのフレームが異なっていれば、現在のフレームの１フレーム後のフレームから比較が再開され、そのフレームはｋ番目のフレームとなる（Ｓ５０４およびＳ５０５）。ｋがｊより小さいか否かがチェックされる（Ｓ５０６）。ｋ番目のフレームがｊ番目のフレームの前であり、ｉ番目のフレームがｋ番目のフレームと類似していれば（Ｓ５１１）、ｋ番目のフレームが現在のフレームとなり（Ｓ５１２）、比較のためのＳ５０２およびＳ５０３の処理が繰り返される。

【００３６】

ｉ番目のフレームがｋ番目のフレームと異なっていれば、ｋが１だけ増加し（Ｓ５０５）、ｋがｊより小さいか否かがチェックされる。ｋがｊより小さくなければ、ｊ番目のフレームはそれ以前のフレームとは異なることを意味し、ｊ番目のフレームは新たな候補テキストチェンジフレームとしてマークされる（Ｓ５０７）。そして、ｊ番目のフレームから新たな検索が始まる（Ｓ５０８）。現在の検索フレームのインデックスｉとＳＴＥＰの和が入力されたビデオフレームの総数ｎＦｒａｍｅより大きければ（Ｓ５０９）、検索は終わり、見つかった候補テキストチェンジフレームは、さらなる処理のために後続部３０２および３０３に送られる。そうでなければ、検索が続けられる。

【００３７】

フレーム間隔ＳＴＥＰの意図は、検索動作全体の時間を短縮することである。ＳＴＥＰが大きすぎれば、ビデオの内容は急に変化し、性能が低下することになる。ＳＴＥＰが小さすぎれば、総検索時間はあまり短縮されないことになる。このフレーム間隔としては、例えば、ＳＴＥＰ＝４フレームが選ばれる。

【００３８】

図１１は、図１０のＳ５０３における２つのフレームの類似度の決定の動作フローチャートを示している。Ｓ５１１における処理のフローチャートは、図１１においてｊをｋに置き換えることで得られる。

【００３９】

開始時に、画像ブロックカウントｎ、有効ブロックカウントｎＶａｌｉｄ、および類似ブロックカウントｎＳｉｍｉｌａｒは、すべて０に設定される（Ｓ５１３）。そして、ｉ番目のフレームとｊ番目のフレームが、Ｎ×Ｎの規模の重複しない小さな画像ブロックに分割され、それらの画像ブロックの数がｎＢｌｏｃｋとして記録される（Ｓ５１４）。例えば、Ｎ＝１６である。２つのフレーム内の同じ位置にある２つの画像ブロックは、画像ブロックペアと定義される。そして、画像ブロックペア毎に、画像ブロック有効化部３１１を用いて、画像ブロックペアが有効ブロックペアであるか否かがチェックされる（Ｓ５１５）。画像ブロックペア毎の変化を検出することにより、２つのフレーム間における変化の検出が達成される。内容が変化したとしても、スライドのバックグラウンド部分は通常は変化しない。したがって、これらの部分の画像ブロックペアは、有効ブロックペアとみなされるべきではない。

【００４０】

ブロックペアが無効であれば、次のブロックペアがチェックされる（Ｓ５１９およびＳ５２０）。ブロックペアが有効ブロックペアであれば、有効ブロックカウントｎＶａｌｉｄが１だけ増加し（Ｓ５１６）、画像ブロック類似度計測部３１２を用いて、２つの画像ブロックの類似度が計測される（Ｓ５１７）。それらのブロックが類似していれば、類似ブロックカウントｎＳｉｍｉｌａｒが１だけ増加する（Ｓ５１８）。すべてのブロックペアが比較されたとき（Ｓ５１９およびＳ５２０）、フレーム類似度判定部３１３を用いて、２つのフレームが類似しているか否かが決定される（Ｓ５２１）。次の条件が満たされれば、２つのフレームは類似しているとみなされる（Ｓ５２２）。

ｎＳｉｍｉｌａｒ＞ｎＶａｌｉｄ＊ｓｉｍｒａｔｅ

例えば、ｓｉｍｒａｔｅ＝０．８５である。上記条件が満たされなければ、２つのフレームは非類似とみなされる（Ｓ５２３）。

【００４１】

図１２は、図１１のＳ５１５における画像ブロック有効化部３１１の動作フロ

ーチャートを示している。まず、ｎ番目の画像ブロックペアの平均と分散が計算される（Ｓ５２４）。ｉ番目のフレームの画像ブロックの濃淡値の平均および分散は、それぞれＭ（ｉ）およびＶ（ｉ）と記される。ｊ番目のフレームの画像ブロックの濃淡値の平均および分散は、それぞれＭ（ｊ）およびＶ（ｊ）と記される。ブロックペアの２つの分散Ｖ（ｉ）およびＶ（ｊ）がどちらも所定のしきい値Ｔｖより小さく（Ｓ５２５）、かつ、２つの平均Ｍ（ｉ）およびＭ（ｊ）の絶対差もまた所定のしきい値Ｔｍより小さければ（Ｓ５２６）、その画像ブロックペアは無効ブロックペアとなる（Ｓ５２７）。そうでなければ、その画像ブロックペアは有効ブロックペアとなる（Ｓ５２８）。

【００４２】

図１３は、図１１のＳ５１７における画像ブロック類似度計測部３１２の動作フローチャートを示している。ｎ番目の画像ブロックペアの平均Ｍ（ｉ）およびＭ（ｊ）がまず計算される（Ｓ５２９）。２つの平均Ｍ（ｉ）およびＭ（ｊ）の絶対差が所定のしきい値Ｔｍ１より大きければ（Ｓ５３０）、それらの２つの画像ブロックは非類似画像ブロックとみなされる（Ｓ５３４）。そうでなければ、２つの画像ブロックの相関Ｃ（ｉ，ｊ）が計算される（Ｓ５３１）。相関Ｃ（ｉ，ｊ）が所定のしきい値Ｔｃより大きければ（Ｓ５３２）、それらの２つの画像ブロックは類似しているとみなされ（Ｓ５３３）、相関Ｃ（ｉ，ｊ）がしきい値Ｔｃより小さければ、それらの２つの画像ブロックは非類似とみなされる（Ｓ５３４）。

【００４３】

図１４から図２１までは、図５のテキストフレーム検出検証部３０２により行われる処理の結果の例を示している。図１４は、元のビデオフレームを示している。図１５は、高速簡易画像二値化の結果である第１の二値画像を示している。図１６は、横二値射影の結果を示している。図１７は、射影正則化の結果を示している。図１８は、候補テキストライン毎の縦二値射影の結果を示している。図１９は、テキストライン領域決定の結果を示している。グレーの矩形は、候補テキストライン領域を示している。

【００４４】

図２０は、図１９において破線によりマークされた２つの候補テキストライン領域の二値画像の２つのペアの結果を示している。最初のペアの二値画像はテキスト情報を含んでいる。これらの２つの画像の違いは非常に小さい。したがって、このテキストライン領域は、真のテキストライン領域とみなされる。２番目のペアの二値画像には非常に大きな違いがある。異なる部分は所定のしきい値より大きいため、この領域は非テキストライン領域とみなされる。図２１は、検出されたテキストライン領域を示している。

【００４５】

図２２および２３は、図３のテキストフレーム検出検証部３０２の動作フローチャートを示している。まず、連続する候補フレームセクションの検出が行われ、フレーム類似度計測部３０１により出力された候補テキストフレームは、一連の連続する候補フレームをそれぞれ含む複数のセクションに分類される（Ｓ７０１）。これらのセクションの数は、ｎＳｅｃｔｉｏｎと記される。最初のセクションから始めて（Ｓ７０２）、ｉ番目のセクションの連続する候補フレームの数Ｍ（ｉ）が所定のしきい値Ｔｎｃｆより大きければ（Ｓ７０３）、高速簡易画像二値化部３２１を用いて、すべてのビデオフレームの各二値画像が求められる（Ｓ７０４）。そして、二値画像の横および縦射影を使用したテキストライン領域決定部３２２を用いて、テキストラインの領域が決定される（Ｓ７０５）。

【００４６】

次に、検出されたテキストライン領域から始めて（Ｓ７０６）、再二値化部３２３を用いて、テキストライン領域の２番目の二値画像が作成される（Ｓ７０７）。再二値化部３２３は、検出されたテキストライン毎の全領域に対してNiblack の画像二値化法を適用して、二値画像を求める。同じテキストライン領域の２つの二値画像は、テキストライン確認部３２４により比較される（Ｓ７０８）。それらの２つの二値画像が類似していれば、ｉ番目のセクションのテキストラインカウントｎＴｅｘｔＬｉｎｅ（ｉ）が１だけ増加する（Ｓ７０９）。この手順はＭ（ｉ）個の連続する候補フレーム内のすべてのテキストラインに対して繰り返される（Ｓ７１０およびＳ７１１）。

【００４７】

非テキストフレームがいくつかのテキストラインを含んでいるとみなされることもあるが、一連の候補フレームがいかなるテキストラインも含んでいないのであれば、これらのフレーム内で検出されるテキストラインの数はそれほど多くならないものと考えられる。そこで、テキストフレーム検証部３２５を用いて、その一連の候補テキストフレームが非テキストフレームであるか否かが確認される。一連の候補テキストフレームは、次の条件が満たされれば非テキストフレームとみなされ（Ｓ７１２）、

$$nTextLine(i) \leq \alpha M(i)$$

これらの偽りの候補テキストフレームは除去される（Ｓ７１３）。ここで、αは、実験により決定される正の実数である。通常は、$\alpha = 0.8$と設定される。この手順は連続する候補フレームのすべてのセクションに対して繰り返される（Ｓ７１４およびＳ７１５）。

【００４８】

図２４は、図２２のＳ７０４における高速簡易画像二値化部３２１の動作フローチャートを示している。フレーム画像は、まずＮ×Ｎの規模の重複しない画像ブロックに分割され、それらの画像ブロックの数はｎＢｌｏｃｋとして記録される（Ｓ７１６）。例えば、Ｎ＝１６である。最初の画像ブロックから始めて（Ｓ７１７）、Niblack の画像二値化法により画像ブロック毎に二値化される（Ｓ７１８）。Niblack の画像二値化のパラメータｋは、ｋ＝－０．４に設定される。この手順はすべての画像ブロックに対して繰り返される（Ｓ７１９およびＳ７２０）。

【００４９】

図２５は、図２４のＳ７１８におけるNiblack の画像二値化法のフローチャートを示している。入力はＭ×Ｎの規模の濃淡画像である。まず、画像の平均Ｍｅａｎおよび分散Ｖａｒが計算される（Ｓ７２１）。分散Ｖａｒが所定のしきい値Ｔｖより小さければ（Ｓ７２２）、その二値画像内のすべての画素が０に設定される。Ｖａｒ＞Ｔｖであれば、次式によりバイナリしきい値Ｔが計算される。

Ｔ＝Ｍｅａｎ＋ｋ＊Ｖａｒ

画像画素ｉ毎に、その画素の濃淡値ｇｒａｙ（ｉ）がＴより大きければ（Ｓ７２６）、その二値画像ｂｉｎ（ｉ）内の画素が０に設定される（Ｓ７２７）、そうでなければ、その画素が１に設定される（Ｓ７２８）。この手順は二値画像内のすべての画素に対して繰り返される（Ｓ７２９およびＳ７３０）。

【００５０】

図２６は、図２２のＳ７０５におけるテキストライン領域決定部３２２の動作フローチャートを示している。このテキストライン領域決定部の入力は、Ｓ７０４からのビデオフレームの二値画像である。横画像射影Ｐｒｊｈがまず計算される（Ｓ７３１）。この射影は平滑化されて（Ｓ７３２）正則化される（Ｓ７３３）。Ｐｒｊｈを正則化した結果はＰｒｊｈｒとなり、０または１の２つの値のみを有する。１はその位置が大きな射影値を有することを意味し、０はその位置が小さな射影値を有することを意味する。Ｐｒｊｈｒにおける各１の領域の始点および終点は、それぞれｓｙ（ｉ）およびｅｙ（ｉ）として記録される（Ｓ７３４）。Ｐｒｊｈｒにおける各１の領域に対して、縦画像射影Ｐｒｊｖ（ｉ）が計算される（Ｓ７３５）。Ｐｒｊｖ（ｉ）は平滑化されて（Ｓ７３６）正則化され、Ｐｒｊｖｒ（ｉ）となる（Ｓ７３７）。Ｐｒｊｖｒ（ｉ）内の２つの１の領域間の距離が２＊領域の高さより小さければ、それらの２つの１の領域は接続されて１つの領域となり、接続後の領域の始点および終点が、それぞれｓｘ（ｉ）およびｅｘ（ｉ）として記録される（Ｓ７３８）。これらの出力ｓｘ（ｉ）、ｅｘ（ｉ）、ｓｙ（ｉ）、およびｅｙ（ｉ）により、ｉ番目のテキストラインの領域が決定される（Ｓ７３９）。

【００５１】

図２７は、図２６のＳ７３１における横画像射影のフローチャートを示している。最初の横ラインから始めて（Ｓ７４０）、ｉ番目の横ラインに対する射影が次式により計算される（Ｓ７４１）。

【００５２】

【数１】

$$prj(i)=\sum_{j=0}^{w-1} I(i,j)$$

【００５３】

ここで、Ｉ（ｉ，ｊ）はｉ行ｊ列の画素値であり、ｗは画像の幅である。この計算は、画像の高さをｈとして、画像内のすべての横ラインに対して繰り返される（Ｓ７４２およびＳ７４３）。

【００５４】

図２８は、図２６のＳ７３２における射影平滑化のフローチャートを示している。平滑化ウィンドウの半径δに相当する点から始めて（Ｓ７４４）、平滑化射影のｉ番目の点の値ｐｒｊｓ（ｉ）が次式により計算される（Ｓ７４５）。

【００５５】

【数２】

$$prjs(i)=\frac{1}{2\delta+1}\sum_{j=i-\delta}^{i+\delta} prj(j)$$

【００５６】

ここで、平滑化ウィンドウの長さを２＊δ＋１とする。この計算は、平滑化の範囲をＬとして、平滑化射影のすべての点に対して繰り返される（Ｓ７４６およびＳ７４７）。

【００５７】

図２９は、図２６のＳ７３３における射影正則化のフローチャートを示している。最初に、射影内のすべての極大値が検出される（Ｓ７４８）。正則化射影Ｐｒｊｒのすべての画素の値は０に設定される（Ｓ７４９）。最初の極大値ｍａｘ（ｉ）から始めて（Ｓ７５０）、すぐ近くの２つの極小値ｍｉｎ１（ｉ）およびｍｉｎ２（ｉ）が検出される（Ｓ７５１）。

【００５８】

図３０は、射影曲線におけるｍａｘ（ｉ）、ｍｉｎ１（ｉ）、およびｍｉｎ２（ｉ）の位置の描画例を示している。３つの極大値が存在している。Ｐ２、Ｐ４、およびＰ６は、それぞれｍａｘ（１）、ｍａｘ（２）、およびｍａｘ（３）と

する。Ｐ１はｍａｘ（１）に対する上側の極小値ｍｉｎ１（１）であり、Ｐ３はｍａｘ（１）に対する下側の極小値ｍｉｎ２（１）である。Ｐ３はｍａｘ（２）に対する上側の極小値ｍｉｎ１（２）でもある。同様に、Ｐ５はｍａｘ（２）に対する下側の極小値ｍｉｎ２（２）であり、ｍａｘ（３）に対する上側の極小値ｍｉｎ１（３）でもある。Ｐ７はｍａｘ（３）に対する下側の極小値ｍｉｎ２（３）である。

【００５９】

ｍｉｎ１（ｉ）＜ｍａｘ（ｉ）／２、かつ、ｍｉｎ２（ｉ）＜ｍａｘ（ｉ）／２であれば（Ｓ７５２）、ｍｉｎ１（ｉ）およびｍｉｎ２（ｉ）の位置の間の正則化射影Ｐｒｊｒの値は１に設定される（Ｓ７５３）。この手順は、極大値毎に繰り返される（Ｓ７５４およびＳ７５５）。

【００６０】

図３１は、図２２のＳ７０８におけるテキストライン確認部３２４の動作フローチャートを示している。このテキストライン確認部の入力は、同じテキストライン領域のｗ×ｈの規模の２つの二値画像Ｉ１およびＩ２である。まず、カウンタｃｏｕｎｔ１、ｃｏｕｎｔ２、およびｃｏｕｎｔが０に設定される（Ｓ７５６）。ｃｏｕｎｔは、Ｉ１およびＩ２内の対応する２つの画素の値がどちらも１であるような画素の数を意味する。ｃｏｕｎｔ１は、Ｉ１内の対応する画素の値が１であり、Ｉ２内のそれが０であるような画素の数を意味する。ｃｏｕｎｔ２は、Ｉ２内の対応する画素の値が１であり、Ｉ１内のそれが０であるような画素の数を意味する。

【００６１】

２つの画像内の最初の位置から始めて、対応する２つの画素Ｉ１（ｉ）およびＩ２（ｉ）が両方とも１であれば、ｃｏｕｎｔが１だけ増加する（Ｓ７５７およびＳ７５８）。そうでない場合、Ｉ１（ｉ）が１であれば、ｃｏｕｎｔ１が１だけ増加する（Ｓ７５９およびＳ７６０）。そうでもない場合、Ｉ２（ｉ）が１であれば、ｃｏｕｎｔ２が１だけ増加する（Ｓ７６１およびＳ７６２）。すべての画素がチェックされた後（Ｓ７６３およびＳ７６４）、次の条件が満たされるか否かがチェックされる（Ｓ７６５およびＳ７６６）。

ｃｏｕｎｔ＋ｃｏｕｎｔ１＜ｗ＊ｈ／２，
ｃｏｕｎｔ＋ｃｏｕｎｔ２＜ｗ＊ｈ／２，
ｃｏｕｎｔ１＜ｃｏｕｎｔ＊０．２，
ｃｏｕｎｔ２＜ｃｏｕｎｔ＊０．２，
ｆｉｌｌｒａｔｅ＜０．５

テキストライン領域の‘ｆｉｌｌｒａｔｅ’は、その領域内の総画素数に対するフォアグランド画素の数の比率として定義される。上記の条件が満たされれば、２つの二値画像はこのテキストライン領域内において類似しているとみなされ、そのテキストライン領域は有効テキストラインとみなされる（Ｓ７６８）。これらの条件の１つでも満たされなければ、そのテキストライン領域は無効テキストラインとみなされる（Ｓ７６７）。

【００６２】

図３２および３３は、図６の画像シフト検出部３０３の動作フローチャートを示している。連続する２つのフレームであるフレームｉおよびｊに対して、まず高速簡易画像二値化部３３１を用いて、それらの２つのフレームの二値画像が求められる（Ｓ８０１）。次に、フレームｉおよびフレームｊに対する横射影Ｐｒｊｙｉおよび Ｐｒｊｙｊをそれぞれ求めるために、テキストライン縦位置決定部３３２を用いて、図２６のＳ７３１のような横画像射影が行われる（Ｓ８０２）。そして、縦シフト検出部３３３を用いて、２つの射影の相関関数Ｃｙ（ｔ）が計算される（Ｓ８０３）。

【００６３】

ここで、２つの射影Ｐｒｊ１（ｘ）およびＰｒｊ２（ｘ）の相関関数Ｃ（ｔ）は、

【００６４】

【数３】

$$C(t)=\frac{1}{L*V1*V2}\sum(\mathrm{Prj1}(x)-\mathrm{M1})*(\mathrm{Prj2}(x+t)-\mathrm{M2})$$

【００６５】

のように定義される。ここで、Ｌは射影の長さであり、Ｍ１およびＭ２はそれぞれ射影Ｐｒｊ１およびＰｒｊ２の平均であり、Ｖ１およびＶ２はそれぞれ射影Ｐｒｊ１およびＰｒｊ２の分散である。

【００６６】

Ｃｙ（ｔ）の最大値が９０％より小さければ（Ｓ８０４）、２つの画像はシフト画像ではない。そうでなければ、Ｃｙ（ｔ）の最大値の位置が縦オフセットｏｆｆｙとして記録され（Ｓ８０５）、射影Ｐｒｊｙｉの正則化射影Ｐｒｊｙｉｒを求めるために、Ｓ７３３のような射影正則化が行われる（Ｓ８０６）。フレームｊがフレームｉのシフトしたものであれば、フレームｊの縦シフトオフセットはｏｆｆｙとなる。Ｐｒｊｙｉｒ内の１の領域はすべて候補テキストライン領域とみなされ、始点ｓｙｉおよび終点ｅｙｉにより表される（Ｓ８０７）。候補テキストライン領域の数はｎＣａｎＴＬとして記録される。

【００６７】

最初の候補テキストライン領域から始めて、一致カウントｎＭａｔｃｈが０に設定される（Ｓ８０８）。フレームｊ内のｃ番目の対応するシフト候補テキストライン領域は、ｓｙｊ（ｃ）＝ｓｙｉ（ｃ）＋ｏｆｆｙおよびｅｙｊ（ｃ）＝ｅｙｉ（ｃ）＋ｏｆｆｙにより表されるものとする（Ｓ８０９）。２つの対応する候補テキストライン領域に対して、縦射影が計算される（Ｓ８１０）。そして、横シフト検出部３３４を用いて、２つの縦射影に対する相関関数Ｃｘ（ｔ）が計算され、Ｃｘ（ｔ）の最大値の位置が、これらの２つの射影に対する横オフセットｏｆｆｘとして記録される（Ｓ８１１）。Ｃｙ（ｔ）の最大値が９０％より大きければ（Ｓ８１２）、２つの候補テキストライン領域は一致したシフトテキストライン領域とみなされ、一致カウントｎＭａｔｃｈが１だけ増加する（Ｓ８１３）。すべての候補テキストラインペアがチェックされた後（Ｓ８１４およびＳ８１５）、一致したシフトテキストライン領域の数がテキストライン領域の数の７０％より大きければ（Ｓ８１６）、フレームｊはフレームｉのシフトしたものとみなされる（Ｓ８１７）。そうでなければ、フレームｊはフレームｉのシフトしたフレームではない（Ｓ８１８）。

【００６８】

図３４は、図１のテキスト抽出装置１０５の構成を示している。テキスト抽出装置は、ビデオフレームのエッジ情報を抽出するエッジ画像生成部９０１、エッジ画像を用いて候補文字ストロークのストローク画像を生成するストローク画像生成部９０２、偽りの文字ストロークを除去するストロークフィルタ部９０３、隣同士のストロークを接続してテキストライン領域にするテキストライン領域形成部９０４、テキストライン領域内の偽りの文字ストロークを削除するテキストライン検証部９０５、およびテキストライン領域の最終二値画像を求めるテキストライン二値化部９０６を備える。テキスト抽出装置の出力は、フレーム内のすべてのテキストライン領域の二値画像のリストである。このテキスト抽出装置１０５によれば、できるだけ多くの偽りのストロークが検出されて除去されるので、テキストライン領域を正確に二値化することができる。

【００６９】

図３５は、図３４のエッジ画像生成部９０１の構成を示している。エッジ画像生成部９０１は、エッジ強度計算部９１１、第１のエッジ画像生成部９１２、および第２のエッジ画像生成部９１３を含む。エッジ強度計算部９１１は、Sobelエッジ検出器を用いてビデオフレーム内の画素毎にエッジ強度を計算する。第１のエッジ画像生成部９１２は、画素毎のエッジ強度を所定のエッジしきい値と比較し、エッジ強度がしきい値より大きければ、第１のエッジ画像内の対応する画素の値をあるバイナリ値に設定し、エッジ強度がしきい値より小さければ、対応する画素の値を他のバイナリ値に設定することにより、第１のエッジ画像を生成する。例えば、あるバイナリ値として論理“１”を用いて白画素を表し、他のバイナリ値として論理“０”を用いて黒画素を表してもよい。第２のエッジ画像生成部９１３は、第１のエッジ画像内の上記あるバイナリ値を有する各画素の位置を中心とするウィンドウ内の画素毎のエッジ強度をそのウィンドウ内の画素の平均エッジ強度と比較し、その画素のエッジ強度が平均エッジ強度より大きければ、第２のエッジ画像内の対応する画素の値を上記あるバイナリ値に設定し、その画素のエッジ強度が平均エッジ強度より小さければ、対応する画素の値を上記他のバイナリ値に設定することにより、第２のエッジ画像を生成する。第２のエッ

ジ画像生成には、例えば、３×３の規模の小さなウィンドウが用いられる。

【００７０】

図３６は、図３４のストローク画像生成部９０２の構成を示している。ストローク画像生成部９０２は、局所画像二値化部９２１を含む。局所画像二値化部９２１は、第２のエッジ画像内の上記あるバイナリ値を有する各画素の位置を中心とするウィンドウを用いることにより、Niblack の二値化方法でビデオフレーム内の濃淡画像を二値化して候補文字ストロークの二値画像を求める。局所画像二値化には、例えば、１１×１１の規模のウィンドウが用いられる。

【００７１】

図３７は、図３４のストロークフィルタ部９０３の構成を示している。ストロークフィルタ部９０３は、ストロークエッジ被覆有効化部９３１および大直線検出部９３２を備える。ストロークエッジ被覆有効化部９３１は、第２のエッジ画像内の上記あるバイナリ値を有する画素と候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、重複率が所定のしきい値より大きければそのストロークを有効ストロークと決定し、重複率が所定のしきい値より小さければそのストロークを無効ストロークと決定して、無効ストロークを偽りのストロークとして除去する。大直線検出部９３２は、ストロークの幅と高さを用いて非常に大きなストロークを偽りのストロークとして除去する。

【００７２】

図３８は、図３４のテキストライン領域形成部９０４の構成を示している。テキストライン領域形成部９０４は、ストローク接続チェック部９４１を含む。ストローク接続チェック部９４１は、２つの隣接するストロークの高さの重複率とそれらの２つのストロークの間の距離とを用いて、それらの２つのストロークが接続可能か否かをチェックする。テキストライン領域形成部９０４は、チェック結果を用いて複数のストロークを１つのテキストライン領域に統合する。

【００７３】

図３９は、図３４のテキストライン検証部９０５の構成を示している。テキストライン検証部９０５は、縦偽りストローク検出部９５１、横偽りストローク検出部９５２、およびテキストライン改善部９５３を含む。縦偽りストローク検出

部９５１は、テキストライン領域内のストロークの平均高さより高い高さの各ストロークをチェックし、そのストロークが２つの横テキストライン領域を接続して１つの大テキストライン領域を生成していれば、そのストロークを偽りのストロークとしてマークする。横偽りストローク検出部９５２は、テキストライン領域内のストロークの平均幅により決定されるしきい値より大きな幅の各ストロークをチェックし、そのストロークを含む領域内のストロークの数が所定のしきい値より小さければ、そのストロークを偽りのストロークとしてマークする。テキストライン改善部９５３は、テキストライン領域内で偽りのストロークが検出されれば、そのテキストライン領域内の偽りのストローク以外のストロークを再接続する。

【００７４】

図４０は、図３４のテキストライン二値化部９０６の構成を示している。テキストライン二値化部９０６は、自動サイズ計算部９６１およびブロック画像二値化部９６２を含む。自動サイズ計算部９６１は、二値化用のウィンドウのサイズを決定する。ブロック画像二値化部９６２は、第２のエッジ画像内の上記あるバイナリ値を有する各画素の位置を中心とする上記ウィンドウを用いることにより、Niblack の二値化方法でビデオフレーム内の濃淡画像を二値化する。偽りのストロークを除去した後のこのようなテキストライン二値化によれば、テキストライン領域を正確に二値化することができる。

【００７５】

図４１から図４６までは、テキスト抽出装置のいくつかの結果を示している。図４１は、元のビデオフレームを示している。図４２は、最終エッジ画像（第２のエッジ画像）となるエッジ画像生成の結果を示している。図４３は、ストローク生成の結果を示している。図４４は、ストロークフィルタリングの結果を示している。図４５は、テキストライン形成の結果を示している。図４６は、精錬された最終二値化テキストライン領域の結果を示している。

【００７６】

図４７および４８は、図３５のエッジ画像生成部９０１の動作フローチャートを示している。まず、Ｗ×Ｈの規模の第１のエッジ画像ＥｄｇｅＩｍｇ１の画素

ＥｄｇｅＩｍｇ１（ｉ）のすべての値が０に設定される（Ｓ１１０１）。そして、最初の画素から始めて（Ｓ１１０２）、エッジ強度計算部９１１を用いて、Sobel エッジ検出器でｉ番目の画素のエッジ強度Ｅ（ｉ）が計算される（Ｓ１１０３）。次に、第１のエッジ画像生成部９１２を用いて、ＥｄｇｅＩｍｇ１（ｉ）の値が決定される。エッジ強度が所定のしきい値Ｔｅｄｇｅより大きければ（Ｓ１１０４）、第１のエッジ画像内のこの画素の値は１に設定され、ＥｄｇｅＩｍｇ１（ｉ）＝１となる（Ｓ１１０５）。この手順はすべての画素がチェックされるまで繰り返される（Ｓ１１０６およびＳ１１０７）。

【００７７】

第１のエッジ画像が得られた後、Ｗ×Ｈの規模の第２のエッジ画像ＥｄｇｅＩｍｇ２に対するすべての値ＥｄｇｅＩｍｇ２（ｉ）が０に初期化される（Ｓ１１０８）。最初の画素から走査して（Ｓ１１０９）、第１のエッジ画像内の画素の値が１であれば（Ｓ１１１０）、図４９に示すような画素ｉの近傍１１１６の配置に従って、近傍画素の平均エッジ強度が求められる（Ｓ１１１１）。そして、第２のエッジ画像生成部９１３を用いて、画素のエッジ強度を平均エッジ強度と比較することにより、第２のエッジ画像内のそれらの近傍画素の値を決定する（Ｓ１１１２）。エッジ強度が平均エッジ強度より大きければ、第２のエッジ画像内の画素値が１に設定され、そうでなければ、値は０に設定される。第１のエッジ画像内のすべての画素がチェックされた後（Ｓ１１１３およびＳ１１１４）、第２のエッジ画像が最終エッジ画像ＥｄｇｅＩｍｇとして出力される（Ｓ１１１５）。

【００７８】

図５０は、図４７のＳ１１０３におけるエッジ強度計算部９１１の動作フローチャートを示している。まず、ｉ番目の画素に対して、図４９に示した近傍領域１１１６内で、横および縦エッジ強度Ｅｘ（ｉ）およびＥｙ（ｉ）が次式により求められる（Ｓ１１１７およびＳ１１１８）。

$$Ex(i) = I(d) + 2*I(e) + I(f) - I(b) - 2*I(a) - I(h),$$

Ｅｙ（ｉ）＝Ｉ（ｂ）＋２＊Ｉ（ｃ）＋Ｉ（ｄ）－Ｉ（ｈ）－２＊Ｉ（ｇ）
－Ｉ（ｆ）

ここで、Ｉ（ｘ）はｘ番目の画素（ｘ＝ａ，ｂ，ｃ，ｄ，ｅ，ｆ，ｇ，ｈ）の濃淡値を表す。総エッジ強度Ｅ（ｉ）は、次式により計算される（Ｓ１１１９）。

【００７９】

【数４】

$$E(i)=\sqrt{Ex(i)*Ex(i)+Ey(i)*Ey(i)}$$

【００８０】

図４８のＳ１１１１における画素ｉの平均エッジ強度は、次式により計算される。

Ｍｅｄｇｅ（ｉ）＝（Ｅ（ａ）＋Ｅ（ｂ）＋Ｅ（ｃ）＋Ｅ（ｄ）＋Ｅ（ｅ）
＋Ｅ（ｆ）＋Ｅ（ｇ）＋Ｅ（ｈ）＋Ｅ（ｉ））／９

図５１は、図３６のストローク画像生成部９０２の動作フローチャートを示している。まず、Ｗ×Ｈの規模のストローク画像が０に初期化される（Ｓ１２０１）。そして、局所画像二値化部９２１を用いて、ストローク画像の画素の値が決定される。最初の画素から始めて（Ｓ１２０２）、エッジ画像ＥｄｇｅＩｍｇのｉ番目の画素の値ＥｄｇｅＩｍｇ（ｉ）が１であれば（Ｓ１２０３）、濃淡フレーム画像内にその画素の位置を中心とする１１×１１のウィンドウが設定され、図２５に示したNiblack の二値化法により、ウィンドウ内のストローク画像の画素の値が決定される。エッジ画像内のすべての画素がチェックされた後（Ｓ１２０５およびＳ１２０６）、ストローク画像が生成される。

【００８１】

図５２は、図３７のストロークフィルタ部９０３の動作フローチャートを示している。まず、大直線検出部９３２を用いて、非常に大きなストロークが削除される。最初のストロークから始めて（Ｓ１３０１）、ストロークの幅または高さ

が所定のしきい値ＭＡＸＳＴＲＯＫＥＳＩＺＥを超えていれば（Ｓ１３０２）、このストロークは削除される（Ｓ１３０４）。そうでなければ、ストロークエッジ被覆有効化部９３１を用いて、そのストロークの有効性がチェックされる（Ｓ１３０３）。有効ストロークは候補文字ストロークを意味し、無効ストロークは真の文字ストロークではない。ストロークが無効であれば、それは削除される（Ｓ１３０４）。ストロークの数をｎＳｔｒｏｋｅとして、ストローク画像内のすべてのストロークに対してチェックが繰り返される（Ｓ１３０５およびＳ１３０６）。

【００８２】

図５３は、図５２のＳ１３０３におけるストロークエッジ被覆有効化部９３１の動作フローチャートを示している。まず、ストロークの輪郭Ｃが求められる（Ｓ１３０７）。現在の輪郭点の近傍領域におけるＥｄｇｅＩｍｇの画素値が、最初の輪郭点から（Ｓ１３０８）チェックされる（Ｓ１３０９）。図４９に示したように、点ａから点ｈまでが点ｉの近傍点とみなされる。１の値を有する隣りのエッジ画素があれば、この輪郭点は有効エッジ輪郭点とみなされ、有効エッジ輪郭点のカウントｎＥｄｇｅが１だけ増加する（Ｓ１３１０）。ｎＣｏｎｔｏｕｒを輪郭点の数としてすべての輪郭点がチェックされた後、有効エッジ輪郭点の数が０．８＊ｎＣｏｎｔｏｕｒより大きければ、そのストロークは有効ストローク、つまり、候補文字ストロークとみなされる（Ｓ１３１４）。そうでなければ、そのストロークは無効ストロークである（Ｓ１３１５）。ｎＣｏｎｔｏｕｒに対するｎＥｄｇｅの比率は重複率を表している。

【００８３】

図５４は、図３８のテキストライン領域形成部９０４の動作フローチャートを示している。まず、すべてのストロークの領域が個々のテキストライン領域として設定され、テキストラインの数ｎＴＬがｎＳｔｒｏｋｅに設定される（Ｓ１４０１）。最初のストロークから始めて（Ｓ１４０２）、ストロークｉの次のストロークｊが選択され（Ｓ１４０３）、ストロークｉおよびストロークｊが１つのテキストライン領域に属するか否かチェックされる（Ｓ１４０４）。否であれば、ストローク接続チェック部９４１を用いて、これらの２つのストロークが接続

可能か否かがチェックされる（Ｓ１４０５）。接続可能であれば、ストロークｉが属するテキストラインおよびストロークｊが属するテキストラインの２つのテキストライン内のすべてのストロークが、１つの大きなテキストラインに統合され（Ｓ１４０６）、テキストラインの数が１だけ減少する（Ｓ１４０７）。

【００８４】

ここで、テキストラインは一群の接続可能なストロークであり、すべてのストロークはテキストラインの属性を持っている。ストロークｉがｍ番目のテキストラインに属し、ストロークｊがｎ番目のテキストラインに属し、ストロークｉがストロークｊと接続可能であれば、ｍ番目およびｎ番目のテキストライン内のすべてのストロークの属性がｍに設定される。すべてのストロークペアがチェックされた後（Ｓ１４０８、Ｓ１４０９、Ｓ１４１０、およびＳ１４１１）、ｎＴＬはそのフレーム内のテキストラインの数となる。

【００８５】

図５５は、図５４のＳ１４０５におけるストローク接続チェック部９４１の動作フローチャートを示している。まず、２つのストロークの高さｈ１およびｈ２が求められ、高い方の高さがｍａｘｈとマークされ、低い方の高さがｍｉｎｈとマークされる（Ｓ１４１２）。ストロークｉおよびストロークｊの中心間の横距離が１．５＊ｍａｘｈより大きければ、これらの２つのストロークは接続可能ではない（Ｓ１４１７）。そうでなければ、ストロークｉおよびストロークｊの両方と交点を持つ横ラインの数がｎＯｖｅｒｌａｐとして記録される（Ｓ１４１４）。ｎＯｖｅｒｌａｐが０．５＊ｍｉｎｈより大きければ（Ｓ１４１５）、これらの２つのストロークは接続可能である（Ｓ１４１６）。そうでなければ、これらの２つのストロークは接続可能ではない（Ｓ１４１７）。Ｓ１４１５におけるｍｉｎｈに対するｎＯｖｅｒｌａｐの比率は重複率を表す。

【００８６】

図５６は、図３９のテキストライン検証部９０５の動作フローチャートを示している。まず、修正フラグｍｏｄｆｌａｇが偽に設定される（Ｓ１５０１）。最初のテキストライン領域から始めて（Ｓ１５０２）、ｉ番目のテキストライン領域の高さＨｅｉｇｈｔ（ｉ）が所定のしきい値ＭＩＮＴＬＨＥＩＧＨＴより小さ

ければ（Ｓ１５０３）、このテキストライン領域は削除される（Ｓ１５０４）。そうでなければ、縦偽りストローク検出部９５１および横偽りストローク検出部９５２を用いて、偽りのストロークが検出される（Ｓ１５０５およびＳ１５０６）。偽りのストロークが検出されれば、そのストロークは削除され（Ｓ１５０７）、テキストライン改善部９５３を用いて残されたストロークが再接続され（Ｓ１５０８）、修正フラグが真に設定される（Ｓ１５０９）。テキストライン改善部９５３は、テキストライン領域形成部９０４と同様にして残されたストロークを再接続する。すべてのテキストライン領域がチェックされた後（Ｓ１５１０およびＳ１５１１）、修正フラグが真であれば（Ｓ１５１２）、偽りのストロークが検出されなくなるまで、再度、全体の処理が繰り返される。

【００８７】

図５７は、図５６のＳ１５０５における縦偽りストローク検出部９５１の動作フローチャートを示している。まず、テキストライン領域内のストロークの平均高さが計算される（Ｓ１５１３）。最初のストロークから始めて（Ｓ１５１４）、ストロークｉの高さが平均高さより大きければ（Ｓ１５１５）、マルチテキストライン検出が行われ、ストロークｉの左側の領域内のストロークがチェックされる（Ｓ１５１６）。ストロークｉの左側の領域はテキストライン領域内部の領域であり、この領域の左、上、および下の境界は、それぞれそのテキストライン領域の左、上、および下の境界である。この領域の右の境界はストロークｉの左の境界である。ストロークｉの左側の領域内に２つ以上の非重複横テキストライン領域が存在すれば、ストロークｉは縦偽りストロークとなる（Ｓ１５２０）。

【００８８】

そうでなければ、次に、マルチテキストライン検出が行われ、ストロークｉの右側の領域内のストロークがチェックされる（Ｓ１５１７）。ストロークｉの右側の領域は、ストロークｉの左側の領域と同様にして定義される。ストロークｉの右側の領域内に２つ以上の非重複横テキストライン領域が存在すれば、ストロークｉは縦偽りストロークとなる（Ｓ１５２０）。この手順はテキストライン領域内のすべてのストロークがチェックされるまで繰り返される（Ｓ１５１８およびＳ１５１９）。

【００８９】

図５８は、図５７のＳ１５１６およびＳ１５１７におけるマルチテキストライン検出のフローチャートを示している。まず、テキストライン領域形成部９０４と同様にしてストロークが接続される。テキストライン領域の数ｎＴｅｘｔＬｉｎｅが１より大きければ（Ｓ１５２２）、次の３つの条件が満たされるか否かがチェックされる。

１．２つの非重複横テキストライン領域が存在する（Ｓ１５２３）。

２．一方のテキストライン領域がもう一方のテキストライン領域の上にある（Ｓ１５２４）。

３．各テキストライン領域内のストロークの数が３より大きい（Ｓ１５２５）。

３つの条件がすべて満たされれば、マルチテキストラインが検出されたことになる（Ｓ１５２６）。

【００９０】

図５９は、図５６のＳ１５０６における横偽りストローク検出部９５２の動作フローチャートを示している。まず、テキストライン領域内のストロークの平均幅が計算される（Ｓ１５２７）。最初のストロークから始めて（Ｓ１５２８）、ストロークｉの幅が平均ストローク幅の２．５倍より大きければ（Ｓ１５２９）、検出領域Ｒが設定される（Ｓ１５３０）。Ｒの左の境界Ｒ．ｌｅｆｔおよび右の境界Ｒ．ｒｉｇｈｔは、それぞれストロークｉの左の境界Ｓｔｒｏｋｅ（ｉ）．Ｌｅｆｔおよび右の境界Ｓｔｒｏｋｅ（ｉ）．Ｒｉｇｈｔにより決定される。Ｒの上の境界Ｒ．ｔｏｐおよび下の境界Ｒ．ｂｏｔｔｏｍは、それぞれそのテキストライン領域の上の境界ｔｅｘｔｌｉｎｅ．ｔｏｐおよび下の境界ｔｅｘｔｌｉｎｅ．ｂｏｔｔｏｍにより決定される。検出領域Ｒ内のストロークの数が計算され（Ｓ１５３１）、その数が３以下であれば（Ｓ１５３２）、ストロークｉは横偽りストロークとしてマークされる（Ｓ１５３３）。この手順はテキストライン領域内のすべてのストロークがチェックされるまで繰り返される（Ｓ１５３４およびＳ１５３５）。

【００９１】

図６０および６１は、偽りのストロークの例を示している。図６０のストロー

ク１５４１は縦偽りストロークであり、図６１のストローク１５４２は横偽りストロークである。

【００９２】

図６２は、図４０のテキストライン二値化部９０６の動作フローチャートを示している。まず、自動サイズ計算部９６１を用いて、テキストライン領域の高さＨｅｉｇｈｔに基づき、次の３つの条件を満たすような二値化用のウィンドウのサイズｗｈが決定される（Ｓ１６０１）。

ｗｈ＝Ｈｅｉｇｈｔ／３，
ｗｈ＝ｗｈ＋１　ｉｆ　ｗｈ　ｉｓ　ａｎ　ｅｖｅｎ　ｎｕｍｂｅｒ，
ｗｈ＝５　ｉｆ　ｗｈ＜５

その後、ブロック画像二値化部９６２を用いて、そのテキストライン領域が再二値化される（Ｓ１６０２）。ブロック画像二値化部９６２は、Niblack の二値化法のウィンドウサイズをｗｈに設定し、ストローク画像生成部９０２と同様にしてテキストライン領域を再二値化する。

【００９３】

図１のビデオテキスト処理装置、あるいはテキストチェンジフレーム検出装置１０４およびテキスト抽出装置１０５の各々は、例えば、図６３に示すような情報処理装置（コンピュータ）を用いて構成される。図６３の情報処理装置は、ＣＰＵ（中央処理装置）１７０１、メモリ１７０２、入力装置１７０３、出力装置１７０４、外部記憶装置１７０５、媒体駆動装置１７０６、ネットワーク接続装置１７０７、およびビデオ入力装置１７０８を備え、それらはバス１７０９により互いに接続されている。

【００９４】

メモリ１７０２は、例えば、ＲＯＭ（read only memory）、ＲＡＭ（random access memory）等を含み、処理に用いられるプログラムおよびデータを格納する。ＣＰＵ１７０１は、メモリ１７０２を利用してプログラムを実行することにより、必要な処理を行う。この場合、図３の３０１から３０３までの各部と図３４

の９０１から９０６までの各部は、メモリ１７０２に格納されたプログラムに対応する。

【００９５】

入力装置１７０３は、例えば、キーボード、ポインティングデバイス、タッチパネル等であり、ユーザからの指示や情報の入力に用いられる。出力装置１７０４は、例えば、ディスプレイ、プリンタ、スピーカ等であり、ユーザへの問い合わせや処理結果の出力に用いられる。

【００９６】

外部記憶装置１７０５は、例えば、磁気ディスク装置、光ディスク装置、光磁気ディスク装置、テープ装置等である。情報処理装置は、この外部記憶装置１７０５に、上記プログラムおよびデータを格納しておき、必要に応じて、それらをメモリ１７０２にロードして使用する。外部記憶装置１７０５は、図１の既存のビデオデータ１０１を格納するデータベースとしても用いられる。

【００９７】

媒体駆動装置１７０６は、可搬記録媒体１７１０を駆動し、その記録内容にアクセスする。可搬記録媒体１７１０は、メモリカード、フレキシブルディスク、ＣＤ－ＲＯＭ（compact disk read only memory ）、光ディスク、光磁気ディスク等の任意のコンピュータ読み取り可能な記録媒体である。ユーザは、この可搬記録媒体１７１０に上記プログラムおよびデータを格納しておき、必要に応じて、それらをメモリ１７０２にロードして使用する。

【００９８】

ネットワーク接続装置１７０７は、ＬＡＮ（local area network）、インターネット等の任意の通信ネットワークに接続され、通信中にデータを変換する。情報処理装置は、上記プログラムおよびデータをネットワーク接続装置１７０７を介して受け取り、必要に応じて、それらをメモリ１７０２にロードして使用する。

【００９９】

ビデオ入力装置１７０８は、例えば、図１のＴＶビデオカメラ１０２であり、生のビデオストリームの入力に用いられる。

図６４は、図６３の情報処理装置にプログラムおよびデータを供給することのできるコンピュータ読み取り可能な記録媒体を示している。可搬記録媒体１７１０やサーバ１８０１のデータベース１８０３に格納されたプログラムおよびデータは、情報処理装置１８０２のメモリ１７０２にロードされる。サーバ１８０１は、そのプログラムおよびデータを搬送する搬送信号を生成し、ネットワーク上の任意の伝送媒体を介して情報処理装置１８０２に送信する。ＣＰＵ１７０１は、そのデータを用いてそのプログラムを実行し、必要な処理を行う。

【０１００】

（付記１）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するテキストチェンジフレーム検出装置であって、

前記与えられたビデオフレームから冗長なビデオフレームを除去する第１のフレーム除去手段と、

前記与えられたビデオフレームからテキスト領域を含まないビデオフレームを除去する第２のフレーム除去手段と、

前記与えられたビデオフレームから画像シフトに起因する冗長なビデオフレームを検出して除去する第３のフレーム除去手段と、

残されたビデオフレームを候補テキストチェンジフレームとして出力する出力手段と

を備えることを特徴とするテキストチェンジフレーム検出装置。

【０１０１】

（付記２）　前記第１のフレーム除去手段は、

前記与えられたビデオフレームのうちの２つのビデオフレーム内の同じ位置にある２つの画像ブロックが、画像コンテンツの変化を示す能力のある有効ブロックペアであるか否かを決定する画像ブロック有効化手段と、

前記有効ブロックペアの２つの画像ブロックの類似度を計算して、該２つの画像ブロックが類似しているか否かを決定する画像ブロック類似度計測手段と、

有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて前記２つのビデオフレームが類似しているか否かを決定するフレーム類似度判定手段とを含み、

前記第１のフレーム除去手段は、類似ビデオフレームを冗長なビデオフレームとして除去することを特徴とする付記１記載のテキストチェンジフレーム検出装置。

【０１０２】

（付記３）　前記第２のフレーム除去手段は、

前記与えられたビデオフレームのうちの１つのビデオフレームの第１の二値画像を生成する高速簡易画像二値化手段と、

前記第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決定するテキストライン領域決定手段と、

テキストライン領域毎に第２の二値画像を生成する再二値化手段と、

前記第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の総数に対する該テキストライン領域内のフォアグランド画素の数の充填率とを用いて、テキストライン領域の有効性を決定するテキストライン確認手段と、

１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、該１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームであるか否かを確認するテキストフレーム検証手段とを含むことを特徴とする付記１記載のテキストチェンジフレーム検出装置。

【０１０３】

（付記４）　前記第３のフレーム除去手段は、

前記与えられたビデオフレームのうちの２つのビデオフレームの二値画像を生成する高速簡易画像二値化手段と、

前記２つのビデオフレームの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定するテキストライン縦位置決定手段と、

前記横射影の相関を用いて、前記２つのビデオフレームの間における画像シフトの縦オフセットと、該２つのビデオフレームの縦方向の類似度とを決定する縦シフト検出手段と、

前記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、前記画像シフトの横オフセットと、該２つのビデオフレームの横方向の類似度とを決定する横シフト検出手段とを含み、

前記第３のフレーム除去手段は、類似ビデオフレームを前記画像シフトに起因する冗長なビデオフレームとして除去することを特徴とする付記１記載のテキストチェンジフレーム検出装置。

【０１０４】

（付記５）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するテキストチェンジフレーム検出装置であって、

与えられたビデオフレームのうちの２つのビデオフレーム内の同じ位置にある２つの画像ブロックが、画像コンテンツの変化を示す能力のある有効ブロックペアであるか否かを決定する画像ブロック有効化手段と、

前記有効ブロックペアの２つの画像ブロックの類似度を計算して、該２つの画像ブロックが類似しているか否かを決定する画像ブロック類似度計測手段と、

有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて前記２つのビデオフレームが類似しているか否かを決定するフレーム類似度判定手段と、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する出力手段と

を備えることを特徴とするテキストチェンジフレーム検出装置。

【０１０５】

（付記６）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するテキストチェンジフレーム検出装置であって、

前記与えられたビデオフレームのうちの１つのビデオフレームの第１の二値画像を生成する高速簡易画像二値化手段と、

前記第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決定するテキストライン領域決定手段と、

テキストライン領域毎に第２の二値画像を生成する再二値化手段と、

前記第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の総数に対する該テキストライン領域内のフォアグランド画素の数の充填率とを用いて、テキストライン領域の有効性を決定するテキストライン確認手段と、

１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、該１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームで

あるか否かを確認するテキストフレーム検証手段と、

前記非テキストフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する出力手段と

を備えることを特徴とするテキストチェンジフレーム検出装置。

【０１０６】

（付記７）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するテキストチェンジフレーム検出装置であって、

前記与えられたビデオフレームのうちの２つのビデオフレームの二値画像を生成する高速簡易画像二値化手段と、

前記２つのビデオフレームの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定するテキストライン縦位置決定手段と、

前記横射影の相関を用いて、前記２つのビデオフレームの間における画像シフトの縦オフセットと、該２つのビデオフレームの縦方向の類似度とを決定する縦シフト検出手段と、

前記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、前記画像シフトの横オフセットと、該２つのビデオフレームの横方向の類似度とを決定する横シフト検出手段と、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する出力手段と

を備えることを特徴とするテキストチェンジフレーム検出装置。

【０１０７】

（付記８）　与えられた画像から少なくとも１つのテキストライン領域を抽出するテキスト抽出装置であって、

前記与えられた画像のエッジ情報を生成するエッジ画像生成手段と、

前記エッジ情報を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成するストローク画像生成手段と、

前記エッジ情報を用いて前記二値画像から偽りのストロークを除去するストロークフィルタ手段と、

複数のストロークを１つのテキストライン領域に統合するテキストライン領域

形成手段と、

前記テキストライン領域から偽りの文字ストロークを除去し、該テキストライン領域を改善するテキストライン検証手段と、

前記テキストライン領域の高さを用いて該テキストライン領域を二値化するテキストライン二値化手段と、

前記テキストライン領域の二値画像を出力する出力手段と

を備えることを特徴とするテキスト抽出装置。

【０１０８】

（付記９）　前記エッジ画像生成手段は、

Sobel エッジ検出器を用いて前記与えられた画像内の画素毎のエッジ強度を計算するエッジ強度計算手段と、

画素毎のエッジ強度を所定のエッジしきい値と比較し、該エッジ強度が該しきい値より大きければ、第１のエッジ画像内の対応する画素の値をあるバイナリ値に設定し、該エッジ強度が該しきい値より小さければ、該対応する画素の値を他のバイナリ値に設定することにより、該第１のエッジ画像を生成する第１のエッジ画像生成手段と、

前記第１のエッジ画像内の前記あるバイナリ値を有する各画素の位置を中心とするウィンドウ内の画素毎のエッジ強度を該ウィンドウ内の画素の平均エッジ強度と比較し、該画素のエッジ強度が該平均エッジ強度より大きければ、第２のエッジ画像内の対応する画素の値を該あるバイナリ値に設定し、該画素のエッジ強度が該平均エッジ強度より小さければ、該対応する画素の値を前記他のバイナリ値に設定することにより、該第２のエッジ画像を生成する第２のエッジ画像生成手段とを含むことを特徴とする付記８記載のテキスト抽出装置。

【０１０９】

（付記１０）　前記ストローク画像生成手段は、前記第２のエッジ画像内の前記あるバイナリ値を有する各画素の位置を中心とするウィンドウを用いることにより、Niblack の二値化方法で前記与えられた画像の濃淡画像を二値化して前記候補文字ストロークの二値画像を求める局所画像二値化手段を含むことを特徴とする付記９記載のテキスト抽出装置。

【０１１０】

（付記１１）　前記ストロークフィルタ手段は、

前記第２のエッジ画像内の前記あるバイナリ値を有する画素と前記候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、該重複率が所定のしきい値より大きければ該ストロークを有効ストロークと決定し、該重複率が該所定のしきい値より小さければ該ストロークを無効ストロークと決定して、該無効ストロークを除去するストロークエッジ被覆有効化手段と、

前記ストロークの幅と高さを用いて大きなストロークを除去する大直線検出手段とを含むことを特徴とする付記９記載のテキスト抽出装置。

【０１１１】

（付記１２）　前記テキストライン二値化手段は、

二値化用のウィンドウのサイズを決定する自動サイズ計算手段と、

前記第２のエッジ画像内の前記あるバイナリ値を有する各画素の位置を中心とする前記ウィンドウを用いることにより、Niblack の二値化方法で前記与えられた画像の濃淡画像を二値化するブロック画像二値化手段とを含むことを特徴とする付記９記載のテキスト抽出装置。

【０１１２】

（付記１３）　前記テキストライン領域形成手段は、２つの隣接するストロークの高さの重複率と該２つのストロークの間の距離とを用いて、該２つのストロークが接続可能か否かをチェックするストローク接続チェック手段を含み、該テキストライン領域形成手段は、チェック結果を用いて前記複数のストロークを１つのテキストライン領域に統合することを特徴とする付記８記載のテキスト抽出装置。

【０１１３】

（付記１４）　前記テキストライン検証手段は、

前記テキストライン領域内のストロークの平均高さより高い高さの各ストロークをチェックし、該ストロークが２つの横テキストライン領域を接続して１つの大テキストライン領域を生成していれば、該ストロークを偽りのストロークとしてマークする縦偽りストローク検出手段と、

前記テキストライン領域内のストロークの平均幅により決定されるしきい値より大きな幅の各ストロークをチェックし、該ストロークを含む領域内のストロークの数が所定のしきい値より小さければ、該ストロークを偽りのストロークとしてマークする横偽りストローク検出手段と、

前記テキストライン領域内で偽りのストロークが検出されれば、該テキストライン領域内の該偽りのストローク以外のストロークを再接続するテキストライン改善手段とを含むことを特徴とする付記８記載のテキスト抽出装置。

【０１１４】

（付記１５）　与えられた画像から少なくとも１つのテキストライン領域を抽出するテキスト抽出装置であって、

前記与えられた画像のエッジ画像を生成するエッジ画像生成手段と、

前記エッジ画像を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成するストローク画像生成手段と、

前記エッジ画像内のエッジを表す画素と前記候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、該重複率が所定のしきい値より大きければ該ストロークを有効ストロークと決定し、該重複率が該所定のしきい値より小さければ該ストロークを無効ストロークと決定して、該無効ストロークを除去するストロークフィルタ手段と、

前記候補文字ストロークの二値画像内の残されたストロークの情報を出力する出力手段と

を備えることを特徴とするテキスト抽出装置。

【０１１５】

（付記１６）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムであって、

前記与えられたビデオフレームから冗長なビデオフレームを除去し、

前記与えられたビデオフレームからテキスト領域を含まないビデオフレームを除去し、

前記与えられたビデオフレームから画像シフトに起因する冗長なビデオフレームを検出して除去し、

残されたビデオフレームを候補テキストチェンジフレームとして出力する処理を前記コンピュータに実行させることを特徴とするプログラム。

【０１１６】

（付記１７）　前記与えられたビデオフレームのうちの２つのビデオフレーム内の同じ位置にある２つの画像ブロックが、画像コンテンツの変化を示す能力のある有効ブロックペアであるか否かを決定し、

前記有効ブロックペアの２つの画像ブロックの類似度を計算して、該２つの画像ブロックが類似しているか否かを決定し、

有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて前記２つのビデオフレームが類似しているか否かを決定し、

類似ビデオフレームを冗長なビデオフレームとして除去する処理を前記コンピュータに実行させることを特徴とする付記１６記載のプログラム。

【０１１７】

（付記１８）　前記与えられたビデオフレームのうちの１つのビデオフレームの第１の二値画像を生成し、

前記第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決定し、

テキストライン領域毎に第２の二値画像を生成し、

前記第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の総数に対する該テキストライン領域内のフォアグランド画素の数の充填率とを用いて、テキストライン領域の有効性を決定し、

１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、該１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームであるか否かを確認する処理を前記コンピュータに実行させることを特徴とする付記１６記載のプログラム。

【０１１８】

（付記１９）　前記プログラムは、

前記与えられたビデオフレームのうちの２つのビデオフレームの二値画像を生成し、

前記２つのビデオフレームの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定し、

前記横射影の相関を用いて、前記２つのビデオフレームの間における画像シフトの縦オフセットと、該２つのビデオフレームの縦方向の類似度とを決定し、

前記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、前記画像シフトの横オフセットと、該２つのビデオフレームの横方向の類似度とを決定する処理を前記コンピュータに実行させ、

前記冗長なビデオフレームを検出して除去するステップは、類似ビデオフレームを前記画像シフトに起因する冗長なビデオフレームとして除去することを特徴とする付記１６記載のプログラム。

【０１１９】

（付記２０）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムであって、

与えられたビデオフレームのうちの２つのビデオフレーム内の同じ位置にある２つの画像ブロックが、画像コンテンツの変化を示す能力のある有効ブロックペアであるか否かを決定し、

前記有効ブロックペアの２つの画像ブロックの類似度を計算して、該２つの画像ブロックが類似しているか否かを決定し、

有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて前記２つのビデオフレームが類似しているか否かを決定し、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【０１２０】

（付記２１）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムであって、

前記与えられたビデオフレームのうちの１つのビデオフレームの第１の二値画像を生成し、

前記第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決

定し、

テキストライン領域毎に第２の二値画像を生成し、

前記第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の総数に対する該テキストライン領域内のフォアグランド画素の数の充填率とを用いて、テキストライン領域の有効性を決定し、

１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、該１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームであるか否かを確認し、

前記非テキストフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【０１２１】

（付記２２）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムであって、

前記与えられたビデオフレームのうちの２つのビデオフレームの二値画像を生成し、

前記２つのビデオフレームの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定し、

前記横射影の相関を用いて、前記２つのビデオフレームの間における画像シフトの縦オフセットと、該２つのビデオフレームの縦方向の類似度とを決定し、

前記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、前記画像シフトの横オフセットと、該２つのビデオフレームの横方向の類似度とを決定し、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【０１２２】

（付記２３）　与えられた画像から少なくとも１つのテキストライン領域を抽出するコンピュータのためのプログラムであって、

前記与えられた画像のエッジ情報を生成し、

前記エッジ情報を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ情報を用いて前記二値画像から偽りのストロークを除去し、

複数のストロークを１つのテキストライン領域に統合し、

前記テキストライン領域から偽りの文字ストロークを除去して、該テキストライン領域を改善し、

前記テキストライン領域の高さを用いて該テキストライン領域を二値化し、

前記テキストライン領域の二値画像を出力する

処理を前記コンピュータに実行させることを特徴とするプログラム。

【０１２３】

（付記２４）　Sobel エッジ検出器を用いて前記与えられた画像内の画素毎のエッジ強度を計算し、

画素毎のエッジ強度を所定のエッジしきい値と比較し、該エッジ強度が該しきい値より大きければ、第１のエッジ画像内の対応する画素の値をあるバイナリ値に設定し、該エッジ強度が該しきい値より小さければ、該対応する画素の値を他のバイナリ値に設定することにより、該第１のエッジ画像を生成し、

前記第１のエッジ画像内の前記あるバイナリ値を有する各画素の位置を中心とするウィンドウ内の画素毎のエッジ強度を該ウィンドウ内の画素の平均エッジ強度と比較し、該画素のエッジ強度が該平均エッジ強度より大きければ、第２のエッジ画像内の対応する画素の値を該あるバイナリ値に設定し、該画素のエッジ強度が該平均エッジ強度より小さければ、該対応する画素の値を前記他のバイナリ値に設定することにより、該第２のエッジ画像を生成する処理を前記コンピュータに実行させることを特徴とする付記２３記載のプログラム。

【０１２４】

（付記２５）　前記第２のエッジ画像内の前記あるバイナリ値を有する各画素の位置を中心とするウィンドウを用いることにより、Niblack の二値化方法で前記与えられた画像の濃淡画像を二値化して前記候補文字ストロークの二値画像を求める処理を前記コンピュータに実行させることを特徴とする付記２４記載のプ

ログラム。

【０１２５】

（付記２６）　前記ストロークの幅と高さを用いて大きなストロークを除去し、

前記第２のエッジ画像内の前記あるバイナリ値を有する画素と前記候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、

前記重複率が所定のしきい値より大きければ前記ストロークを有効ストロークと決定し、該重複率が該所定のしきい値より小さければ該ストロークを無効ストロークと決定し、

前記無効ストロークを除去する処理を前記コンピュータに実行させることを特徴とする付記２４記載のプログラム。

【０１２６】

（付記２７）　二値化用のウィンドウのサイズを決定し、

前記第２のエッジ画像内の前記あるバイナリ値を有する各画素の位置を中心とする前記ウィンドウを用いることにより、Niblack の二値化方法で前記与えられた画像の濃淡画像を二値化する処理を前記コンピュータに実行させることを特徴とする付記２４記載のプログラム。

【０１２７】

（付記２８）　２つの隣接するストロークの高さの重複率と該２つのストロークの間の距離とを用いて、該２つのストロークが接続可能か否かをチェックし、チェック結果を用いて前記複数のストロークを１つのテキストライン領域に統合する処理を前記コンピュータに実行させることを特徴とする付記２３記載のプログラム。

【０１２８】

（付記２９）　前記テキストライン領域内のストロークの平均高さより高い高さの各ストロークをチェックし、

前記ストロークが２つの横テキストライン領域を接続して１つの大テキストライン領域を生成していれば、該ストロークを偽りのストロークとしてマークし、

前記テキストライン領域内のストロークの平均幅により決定されるしきい値よ

り大きな幅の各ストロークをチェックし、

前記ストロークを含む領域内のストロークの数が所定のしきい値より小さければ、該ストロークを偽りのストロークとしてマークし、

前記テキストライン領域内で偽りのストロークが検出されれば、該テキストライン領域内の該偽りのストローク以外のストロークを再接続する処理を前記コンピュータに実行させることを特徴とする付記２３記載のプログラム。

【０１２９】

（付記３０）　与えられた画像から少なくとも１つのテキストライン領域を抽出するコンピュータのためのプログラムであって、

前記与えられた画像のエッジ画像を生成し、

前記エッジ画像を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ画像内のエッジを表す画素と前記候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、

前記重複率が所定のしきい値より大きければ該ストロークを有効ストロークと決定し、該重複率が該所定のしきい値より小さければ該ストロークを無効ストロークと決定し、

前記無効ストロークを除去し、

前記候補文字ストロークの二値画像内の残されたストロークの情報を出力する処理を前記コンピュータに実行させることを特徴とするプログラム。

【０１３０】

（付記３１）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムを記録した記録媒体であって、該プログラムは、

前記与えられたビデオフレームから冗長なビデオフレームを除去し、

前記与えられたビデオフレームからテキスト領域を含まないビデオフレームを除去し、

前記与えられたビデオフレームから画像シフトに起因する冗長なビデオフレームを検出して除去し、

残されたビデオフレームを候補テキストチェンジフレームとして出力する
処理を前記コンピュータに実行させることを特徴とするコンピュータ読み取り可能な記録媒体。

【０１３１】

（付記３２）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムを記録した記録媒体であって、該プログラムは、

与えられたビデオフレームのうちの２つのビデオフレーム内の同じ位置にある２つの画像ブロックが、画像コンテンツの変化を示す能力のある有効ブロックペアであるか否かを決定し、

前記有効ブロックペアの２つの画像ブロックの類似度を計算して、該２つの画像ブロックが類似しているか否かを決定し、

有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて前記２つのビデオフレームが類似しているか否かを決定し、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する
処理を前記コンピュータに実行させることを特徴とするコンピュータ読み取り可能な記録媒体。

【０１３２】

（付記３３）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムを記録した記録媒体であって、該プログラムは、

前記与えられたビデオフレームのうちの１つのビデオフレームの第１の二値画像を生成し、

前記第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決定し、

テキストライン領域毎に第２の二値画像を生成し、

前記第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の総数に対する該テキストライン領域内のフォアグランド画素の数の充填率とを用

いて、テキストライン領域の有効性を決定し、

１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、該１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームであるか否かを確認し、

前記非テキストフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするコンピュータ読み取り可能な記録媒体。

【０１３３】

（付記３４）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータのためのプログラムを記録した記録媒体であって、該プログラムは、

前記与えられたビデオフレームのうちの２つのビデオフレームの二値画像を生成し、

前記２つのビデオフレームの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定し、

前記横射影の相関を用いて、前記２つのビデオフレームの間における画像シフトの縦オフセットと、該２つのビデオフレームの縦方向の類似度とを決定し、

前記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、前記画像シフトの横オフセットと、該２つのビデオフレームの横方向の類似度とを決定し、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とするコンピュータ読み取り可能な記録媒体。

【０１３４】

（付記３５）　与えられた画像から少なくとも１つのテキストライン領域を抽出するコンピュータのためのプログラムを記録した記録媒体であって、該プログラムは、

前記与えられた画像のエッジ情報を生成し、

前記エッジ情報を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ情報を用いて前記二値画像から偽りのストロークを除去し、

複数のストロークを１つのテキストライン領域に統合し、

前記テキストライン領域から偽りの文字ストロークを除去して、該テキストライン領域を改善し、

前記テキストライン領域の高さを用いて該テキストライン領域を二値化し、

前記テキストライン領域の二値画像を出力する

処理を前記コンピュータに実行させることを特徴とするコンピュータ読み取り可能な記録媒体。

【０１３５】

（付記３６）　与えられた画像から少なくとも１つのテキストライン領域を抽出するコンピュータのためのプログラムを記録した記録媒体であって、該プログラムは、

前記与えられた画像のエッジ画像を生成し、

前記エッジ画像を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ画像内のエッジを表す画素と前記候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、

前記重複率が所定のしきい値より大きければ該ストロークを有効ストロークと決定し、該重複率が該所定のしきい値より小さければ該ストロークを無効ストロークと決定し、

前記無効ストロークを除去し、

前記候補文字ストロークの二値画像内の残されたストロークの情報を出力する処理を前記コンピュータに実行させることを特徴とするコンピュータ読み取り可能な記録媒体。

【０１３６】

（付記３７）　与えられたビデオフレームからテキストコンテンツを含む複数

のビデオフレームを選択するコンピュータにプログラムを搬送する搬送信号であって、該プログラムは、

前記与えられたビデオフレームから冗長なビデオフレームを除去し、

前記与えられたビデオフレームからテキスト領域を含まないビデオフレームを除去し、

前記与えられたビデオフレームから画像シフトに起因する冗長なビデオフレームを検出して除去し、

残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とする搬送信号。

【０１３７】

（付記３８）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータにプログラムを搬送する搬送信号であって、該プログラムは、

与えられたビデオフレームのうちの２つのビデオフレーム内の同じ位置にある２つの画像ブロックが、画像コンテンツの変化を示す能力のある有効ブロックペアであるか否かを決定し、

前記有効ブロックペアの２つの画像ブロックの類似度を計算して、該２つの画像ブロックが類似しているか否かを決定し、

有効ブロックペアの総数に対する類似画像ブロックの数の比を用いて前記２つのビデオフレームが類似しているか否かを決定し、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とする搬送信号。

【０１３８】

（付記３９）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータにプログラムを搬送する搬送信号であって、該プログラムは、

前記与えられたビデオフレームのうちの１つのビデオフレームの第１の二値画像を生成し、

前記第１の二値画像の横射影と縦射影を用いてテキストライン領域の位置を決定し、

テキストライン領域毎に第２の二値画像を生成し、

前記第１の二値画像と第２の二値画像の差と、テキストライン領域内の画素の総数に対する該テキストライン領域内のフォアグランド画素の数の充填率とを用いて、テキストライン領域の有効性を決定し、

１組の連続するビデオフレーム内の有効テキストライン領域の数を用いて、該１組の連続するビデオフレームがテキスト領域を含まない非テキストフレームであるか否かを確認し、

前記非テキストフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とする搬送信号。

【０１３９】

（付記４０）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するコンピュータにプログラムを搬送する搬送信号であって、該プログラムは、

前記与えられたビデオフレームのうちの２つのビデオフレームの二値画像を生成し、

前記２つのビデオフレームの二値画像の横射影を用いてテキストライン領域毎の縦位置を決定し、

前記横射影の相関を用いて、前記２つのビデオフレームの間における画像シフトの縦オフセットと、該２つのビデオフレームの縦方向の類似度とを決定し、

前記２つのビデオフレームの二値画像内におけるテキストライン毎の縦射影の相関を用いて、前記画像シフトの横オフセットと、該２つのビデオフレームの横方向の類似度とを決定し、

類似ビデオフレームを除去した後に残されたビデオフレームを候補テキストチェンジフレームとして出力する

処理を前記コンピュータに実行させることを特徴とする搬送信号。

【０１４０】

（付記４１）　与えられた画像から少なくとも１つのテキストライン領域を抽出するコンピュータにプログラムを搬送する搬送信号であって、該プログラムは、

前記与えられた画像のエッジ情報を生成し、

前記エッジ情報を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ情報を用いて前記二値画像から偽りのストロークを除去し、

複数のストロークを１つのテキストライン領域に統合し、

前記テキストライン領域から偽りの文字ストロークを除去して、該テキストライン領域を改善し、

前記テキストライン領域の高さを用いて該テキストライン領域を二値化し、

前記テキストライン領域の二値画像を出力する

処理を前記コンピュータに実行させることを特徴とする搬送信号。

【０１４１】

（付記４２）　与えられた画像から少なくとも１つのテキストライン領域を抽出するコンピュータにプログラムを搬送する搬送信号であって、該プログラムは、

前記与えられた画像のエッジ画像を生成し、

前記エッジ画像を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ画像内のエッジを表す画素と前記候補文字ストロークの二値画像内のストロークの輪郭との重複率をチェックし、

前記重複率が所定のしきい値より大きければ該ストロークを有効ストロークと決定し、該重複率が該所定のしきい値より小さければ該ストロークを無効ストロークと決定し、

前記無効ストロークを除去し、

前記候補文字ストロークの二値画像内の残されたストロークの情報を出力する

処理を前記コンピュータに実行させることを特徴とする搬送信号。

【０１４２】

（付記４３）　与えられたビデオフレームからテキストコンテンツを含む複数のビデオフレームを選択するテキストチェンジフレーム検出方法であって、

前記与えられたビデオフレームから冗長なビデオフレームを除去し、

前記与えられたビデオフレームからテキスト領域を含まないビデオフレームを除去し、

前記与えられたビデオフレームから画像シフトに起因する冗長なビデオフレームを検出して除去し、

残されたビデオフレームを候補テキストチェンジフレームとして提示する

ことを特徴とするテキストチェンジフレーム検出方法。

【０１４３】

（付記４４）　与えられた画像から少なくとも１つのテキストライン領域を抽出するテキスト抽出方法であって、

前記与えられた画像のエッジ情報を生成し、

前記エッジ情報を用いて前記与えられた画像内の候補文字ストロークの二値画像を生成し、

前記エッジ情報を用いて前記二値画像から偽りのストロークを除去し、

複数のストロークを１つのテキストライン領域に統合し、

前記テキストライン領域から偽りの文字ストロークを除去して、該テキストライン領域を改善し、

前記テキストライン領域の高さを用いて該テキストライン領域を二値化し、

前記テキストライン領域の二値画像を提示する

ことを特徴とするテキスト抽出方法。

【０１４４】

【発明の効果】

本発明によれば、テキスト領域を含んでいないビデオフレームとともに、重複しているビデオフレームおよびシフトしているビデオフレームを、与えられたビデオフレームから非常に高速に除去することができる。さらに、できるだけ多くの偽りのストロークが検出されて除去されるので、ビデオフレーム内のテキストライン領域を正確に二値化することができる。

【図面の簡単な説明】

【図１】

本発明のビデオテキスト処理装置の構成を示す図である。

【図２】

ビデオテキスト処理装置の処理フローチャートである。

【図３】

本発明のテキストチェンジフレーム検出装置の構成を示すブロック図である。

【図４】

フレーム類似度計測部の構成を示す図である。

【図５】

テキストフレーム検出検証部の構成を示す図である。

【図６】

画像シフト検出部の構成を示す図である。

【図７】

テキストコンテンツを有する第１のフレームを示す図である。

【図８】

テキストコンテンツを有する第２のフレームを示す図である。

【図９】

フレーム類似度計測部の処理結果を示す図である。

【図１０】

フレーム類似度計測部の動作フローチャートである。

【図１１】

２つのフレームの類似度の決定のフローチャートである。

【図１２】

画像ブロック有効化部の動作フローチャートである。

【図１３】

画像ブロック類似度計測部の動作フローチャートである。

【図１４】

テキストフレーム検出検証のための元のビデオフレームを示す図である。

【図１５】

高速簡易画像二値化の結果の第１の二値画像を示す図である。

【図１６】

横射影の結果を示す図である。

【図１７】

射影正則化の結果を示す図である。

【図１８】

候補テキストライン毎の縦二値射影の結果を示す図である。

【図１９】

テキストライン領域決定の結果を示す図である。

【図２０】

２つの候補テキストラインに対する２組の二値画像を示す図である。

【図２１】

検出されたテキストライン領域を示す図である。

【図２２】

テキストフレーム検出検証部の動作フローチャート（その１）である。

【図２３】

テキストフレーム検出検証部の動作フローチャート（その２）である。

【図２４】

高速簡易二値化部の動作フローチャートである。

【図２５】

Niblack の画像二値化法のフローチャートである。

【図２６】

テキストライン領域決定部の動作フローチャートである。

【図２７】

横画像射影のフローチャートである。

【図２８】

射影平滑化のフローチャートである。

【図２９】

射影正則化のフローチャートである。

【図３０】

射影のｍａｘおよびｍｉｎの例を示す図である。

【図３１】

テキストライン確認部の動作フローチャートである。

【図３２】

画像シフト検出部の動作フローチャート（その１）である。

【図３３】

画像シフト検出部の動作フローチャート（その２）である。

【図３４】

本発明のテキスト抽出装置の構成を示す図である。

【図３５】

エッジ画像生成部の構成を示す図である。

【図３６】

ストローク画像生成部の構成を示す図である。

【図３７】

ストロークフィルタ部の構成を示す図である。

【図３８】

テキストライン領域形成部の構成を示す図である。

【図３９】

テキストライン検証部の構成を示す図である。

【図４０】

テキストライン二値化部の構成を示す図である。

【図４１】

テキスト抽出のための元のビデオフレームを示す図である。

【図４２】

エッジ画像生成の結果を示す図である。

【図４３】

ストローク生成の結果を示す図である。

【図４４】

ストロークフィルタリングの結果を示す図である。

【図４５】

テキストライン領域形成の結果を示す図である。

【図４６】

最終二値化テキストライン領域を示す図である。

【図４７】

エッジ画像生成部の動作フローチャート（その１）である。

【図４８】

エッジ画像生成部の動作フローチャート（その２）である。

【図４９】

画素ｉの近傍の配置を示す図である。

【図５０】

エッジ強度計算部の動作フローチャートである。

【図５１】

ストローク画像生成部の動作フローチャートである。

【図５２】

ストロークフィルタ部の動作フローチャートである。

【図５３】

ストロークエッジ被覆有効化部の動作フローチャートである。

【図５４】

テキストライン領域形成部の動作フローチャートである。

【図５５】

ストローク接続チェック部の動作フローチャートである。

【図５６】

テキストライン検証部の動作フローチャートである。

【図５７】

縦偽りストローク検出部の動作フローチャートである。

【図５８】

マルチテキストライン検出のフローチャートである。

【図５９】

横偽りストローク検出部の動作フローチャートである。

【図６０】

第１の偽りストロークを示す図である。

【図６１】

第２の偽りストロークを示す図である。

【図６２】

テキストライン二値化部の動作フローチャートである。

【図６３】

情報処理装置の構成を示す図である。

【図６４】

記録媒体を示す図である。

【符号の説明】

１０１　ビデオデータ

１０２　ＴＶビデオカメラ

１０３　ビデオ分解部

１０４　テキストチェンジフレーム検出装置

１０５　テキスト抽出装置

１０６、１８０３　データベース

３０１　フレーム類似度計測部

３０２　テキストフレーム検出検証部

３０３　画像シフト検出部

３１１　画像ブロック有効化部

３１２　画像ブロック類似度計測部

３１３　フレーム類似度判定部

３２１、３３１　高速簡易画像二値化部

３２２　テキストライン領域決定部

３２３　再二値化部

３２４　テキストライン確認部

３２５　テキストフレーム検証部

３３２　テキストライン縦位置決定部

３３３　縦シフト検出部

３３４　横シフト検出部

９０１　エッジ画像生成部

９０２　ストローク画像生成部

９０３　ストロークフィルタ部

９０４　テキストライン領域形成部

９０５　テキストライン検証部

９０６　テキストライン二値化部

９１１　エッジ強度計算部

９１２　第１のエッジ画像生成部

９１３　第２のエッジ画像生成部

９２１　局所画像二値化部

９３１　ストロークエッジ被覆有効化部

９３２　大直線検出部

９４１　ストローク接続チェック部

９５１　縦偽りストローク検出部

９５２　横偽りストローク検出部

９５３　テキストライン改善部

９６１　自動サイズ計算部

９６２　画像ブロック二値化部

１５４１、１５４２　偽りストローク

１７０１　ＣＰＵ

１７０２　メモリ

１７０３　入力装置

１７０４　出力装置

１７０５　外部記憶装置

１７０６　媒体駆動装置

１７０７　ネットワーク接続装置

１７０８　ビデオ入力装置

１７０９　バス

１７１０　可搬記録媒体

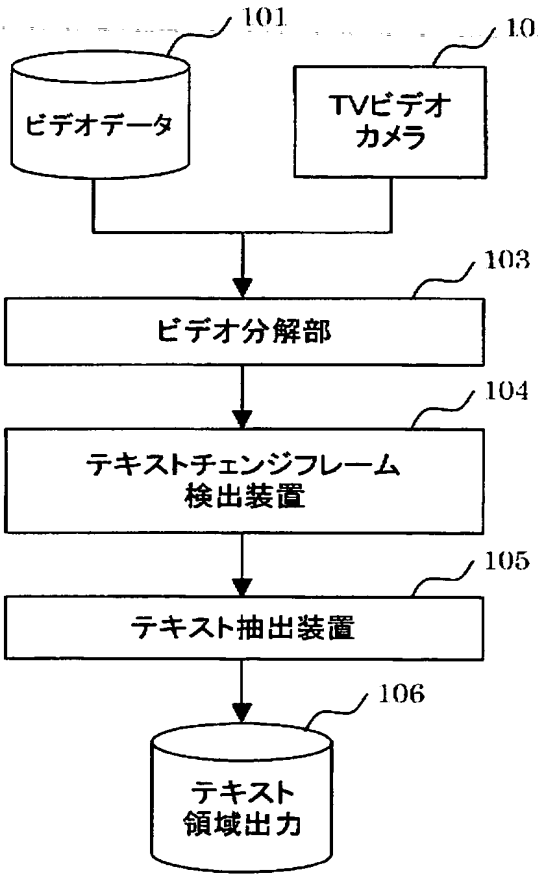
１８０１　サーバ

１８０２　情報処理装置

【書類名】　図面

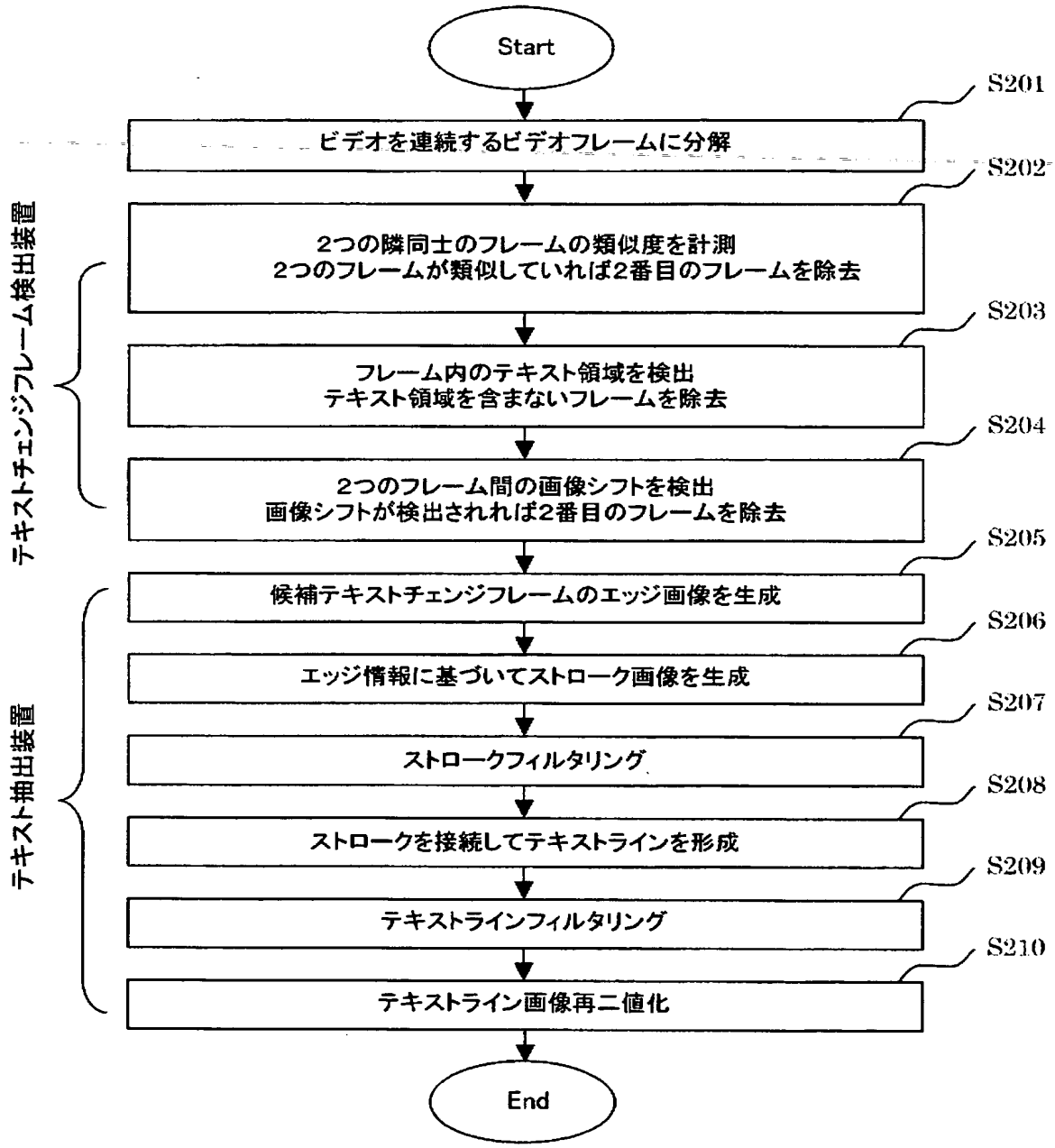
【図１】

本発明のビデオテキスト処理装置の構成を示す図



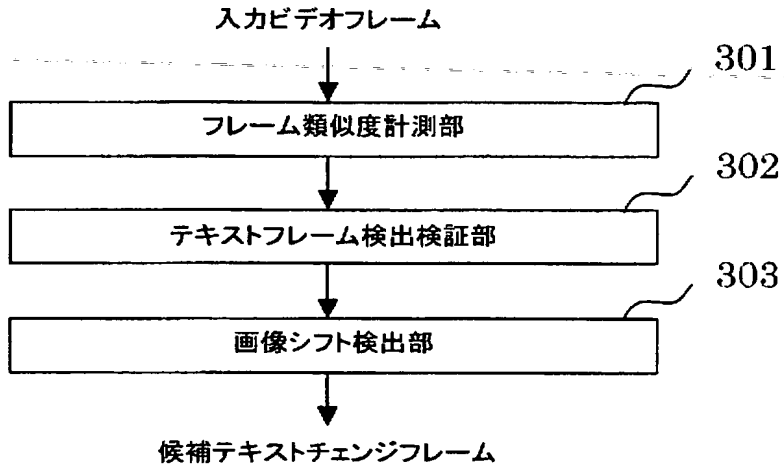
【図２】

ビデオテキスト処理装置の処理フローチャート



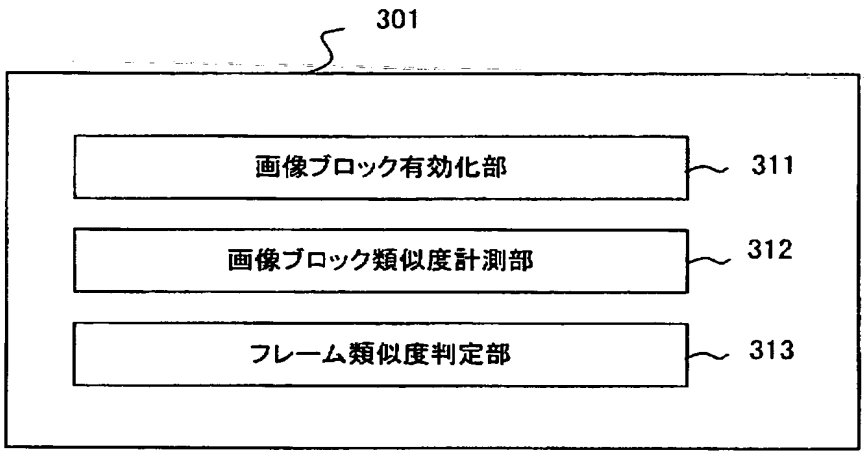
【図３】

本発明のテキストチェンジフレーム検出装置の構成を示すブロック図



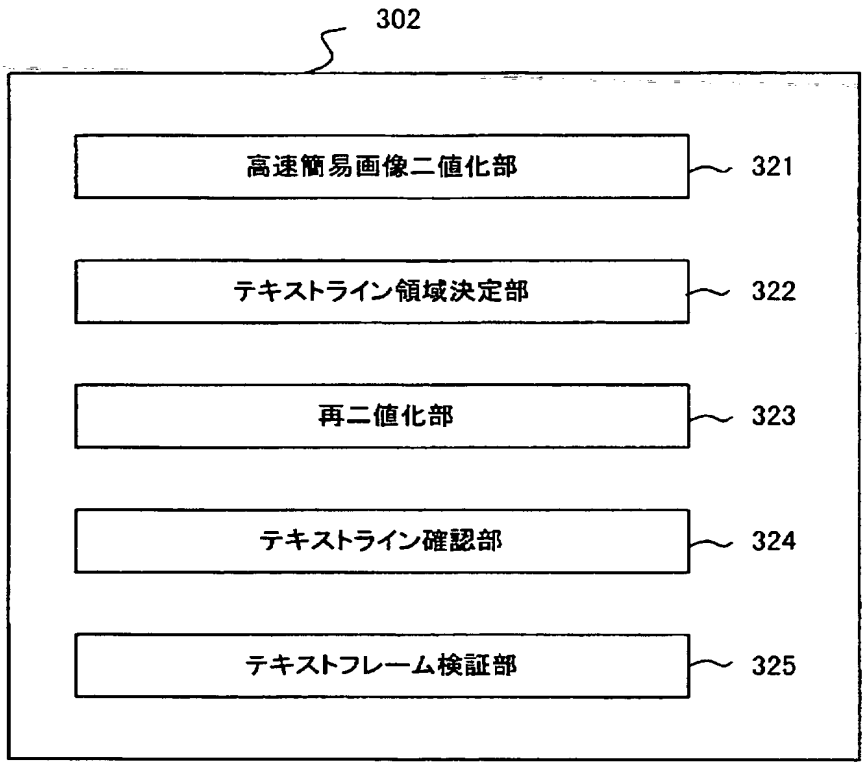
【図４】

フレーム類似度計測部の構成を示す図



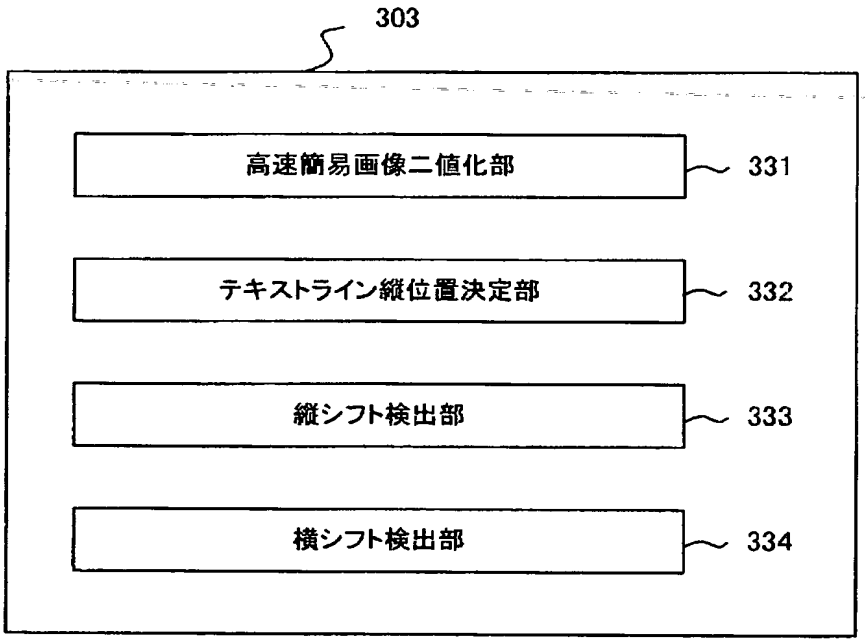
【図５】

テキストフレーム検出検証部の構成を示す図



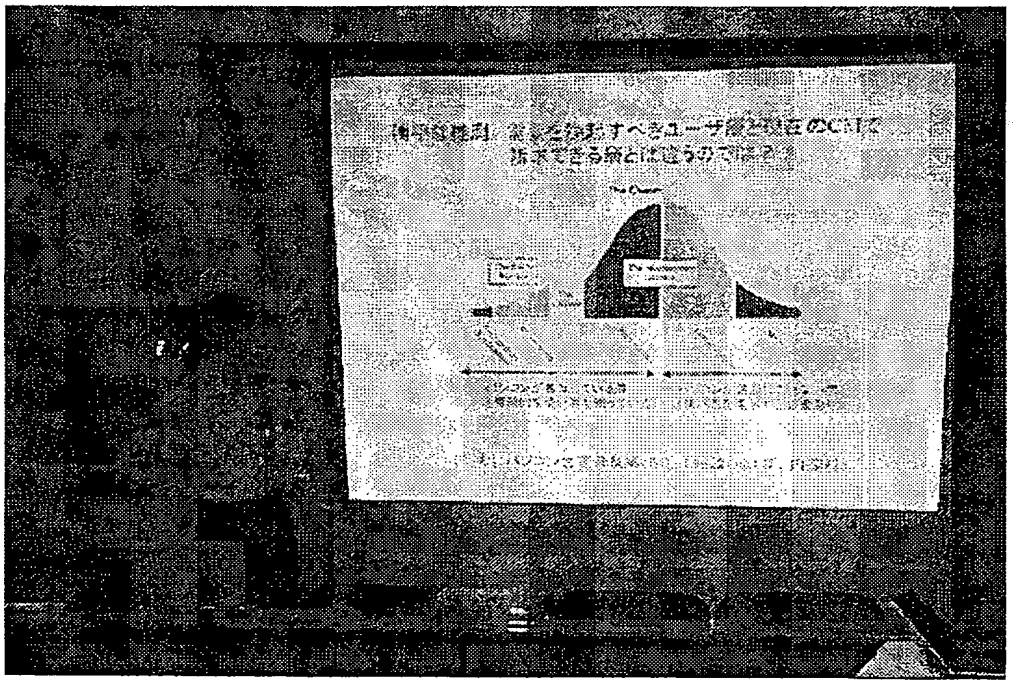
【図６】

画像シフト検出部の構成を示す図



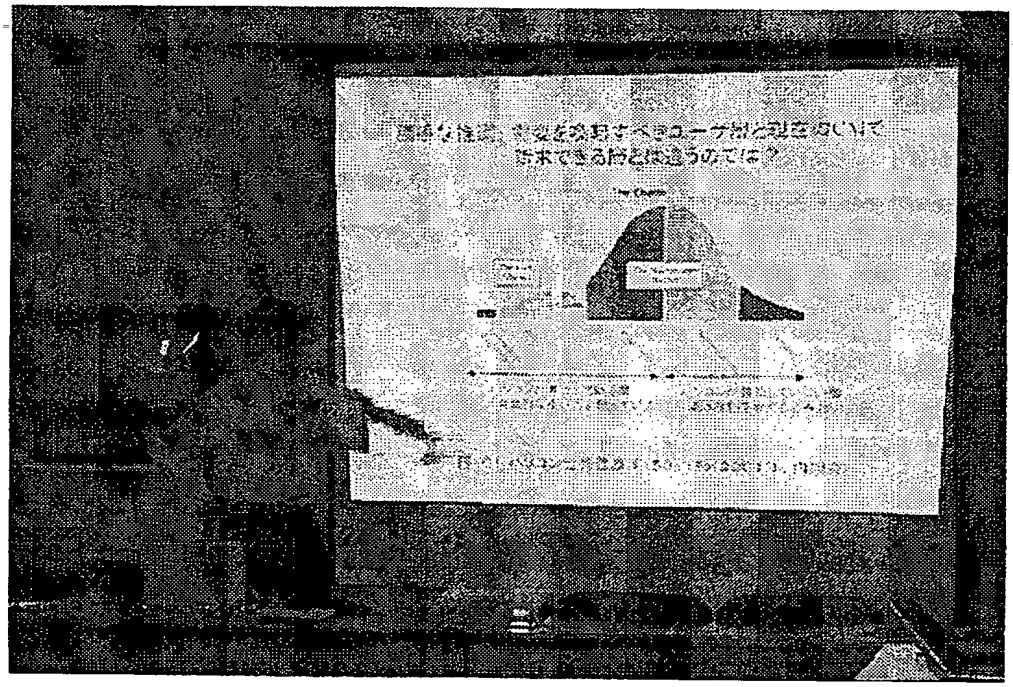
【図７】

テキストコンテンツを有する第１のフレームを示す図



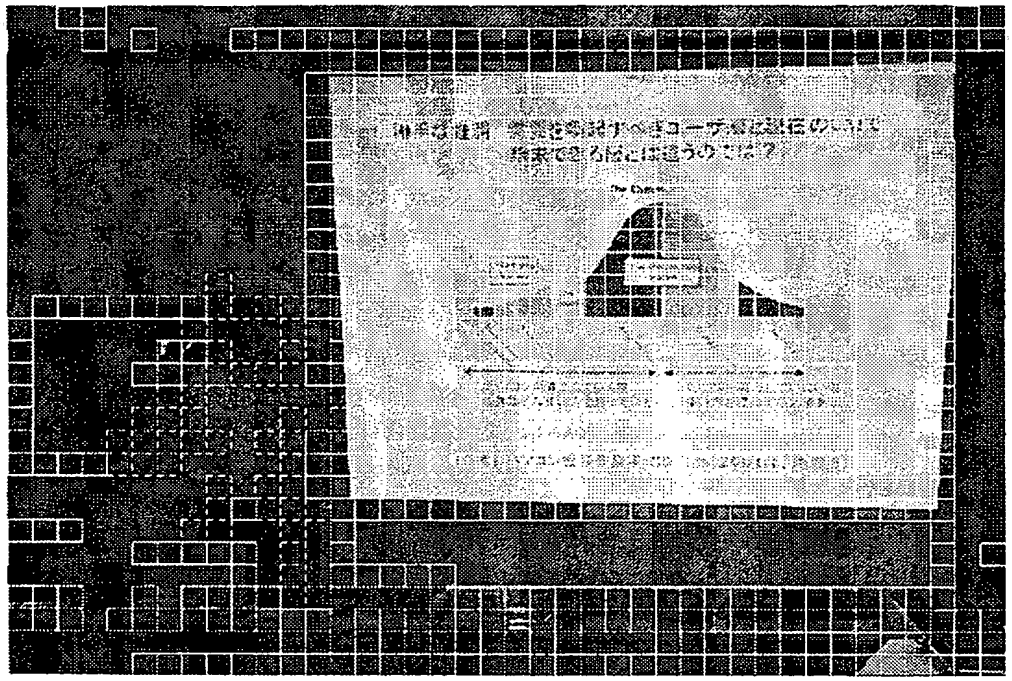
【図８】

テキストコンテンツを有する第2のフレームを示す図



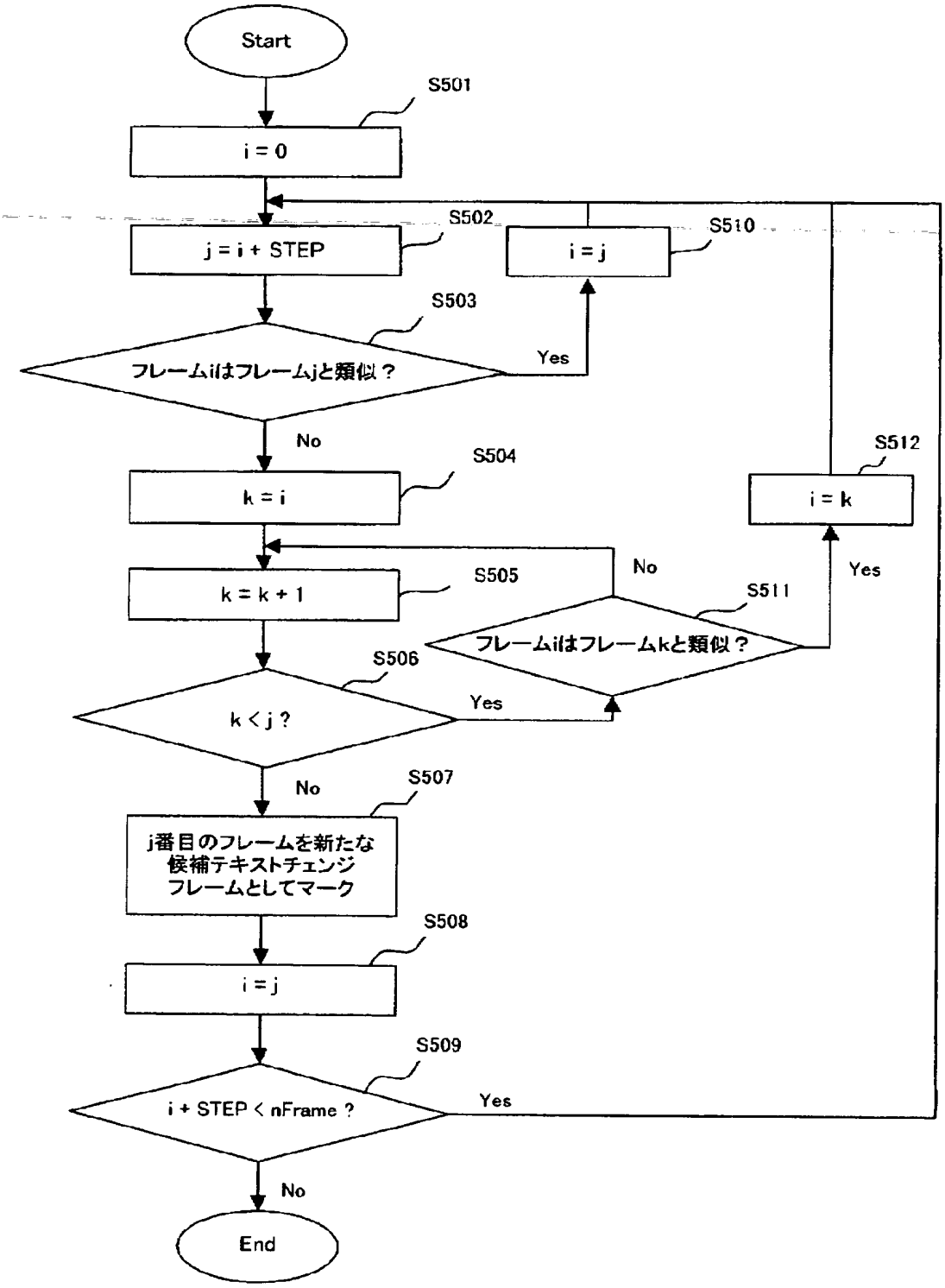
【図９】

フレーム類似度計測部の処理結果を示す図



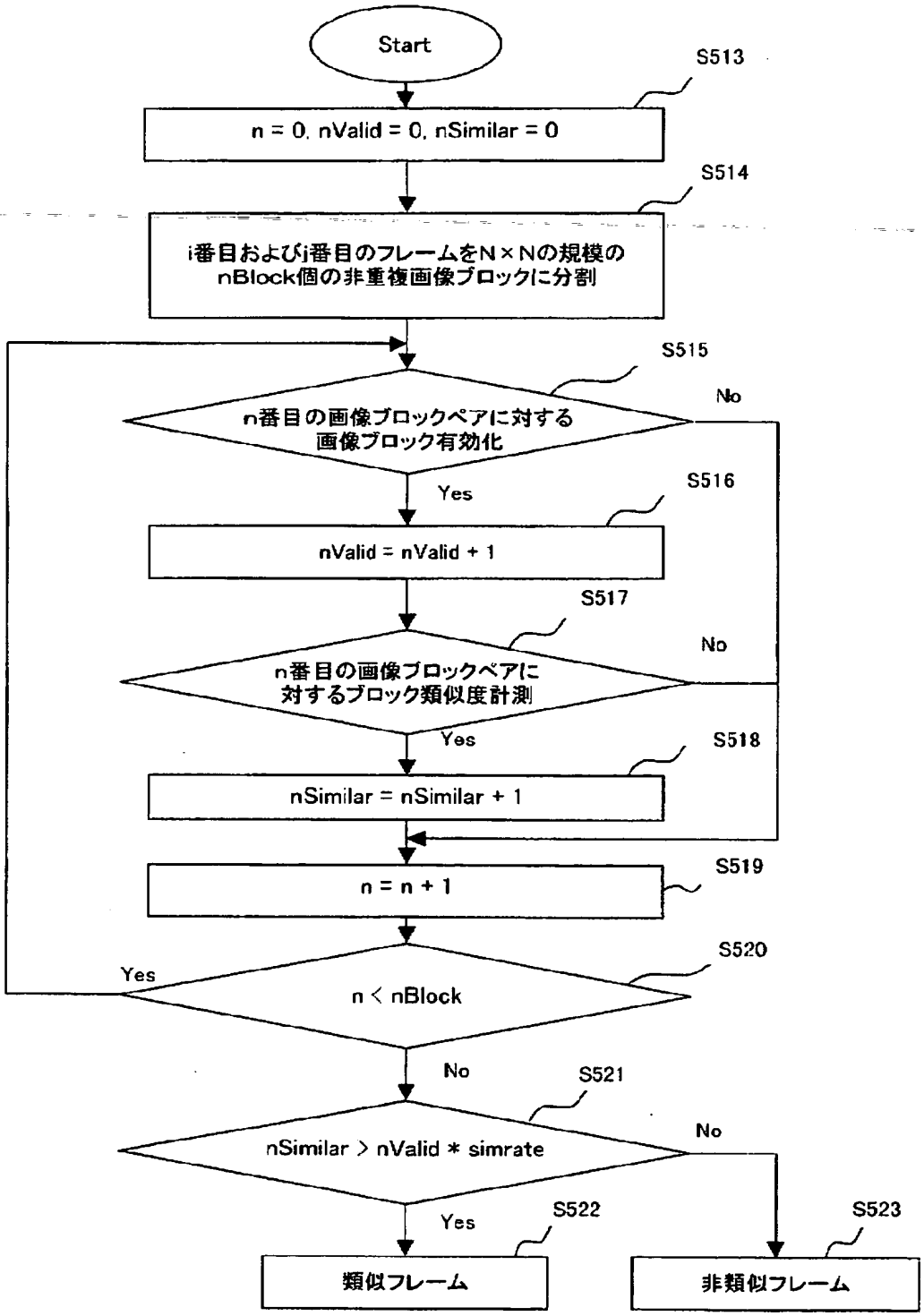
【図１０】

フレーム類似度計測部の動作フローチャート



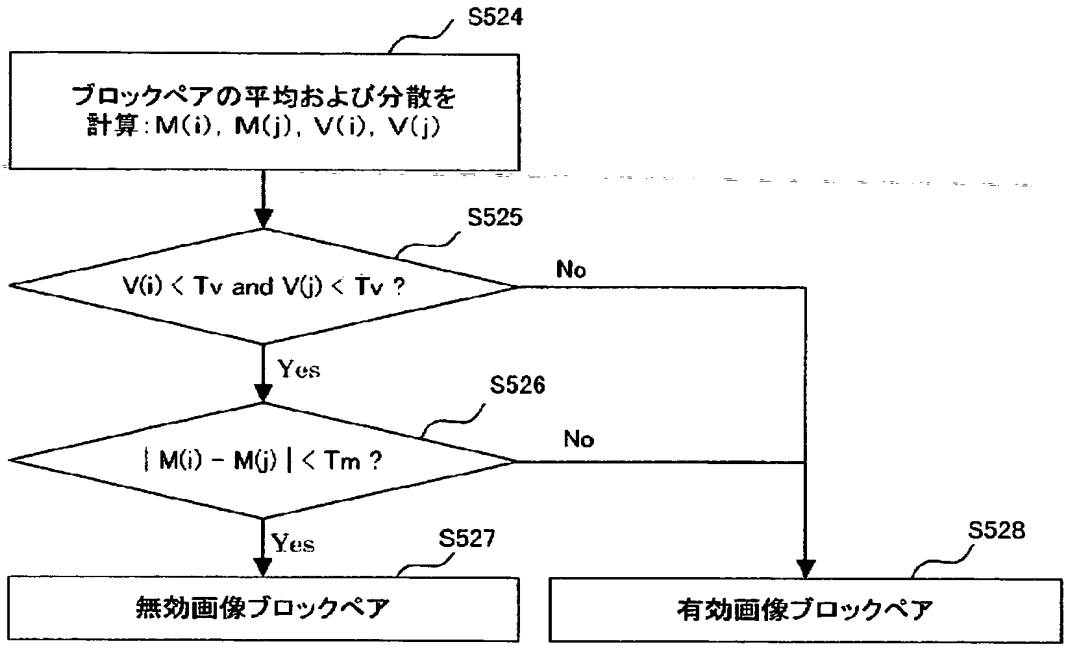
【図１１】

2つのフレームの類似度の決定のフローチャート



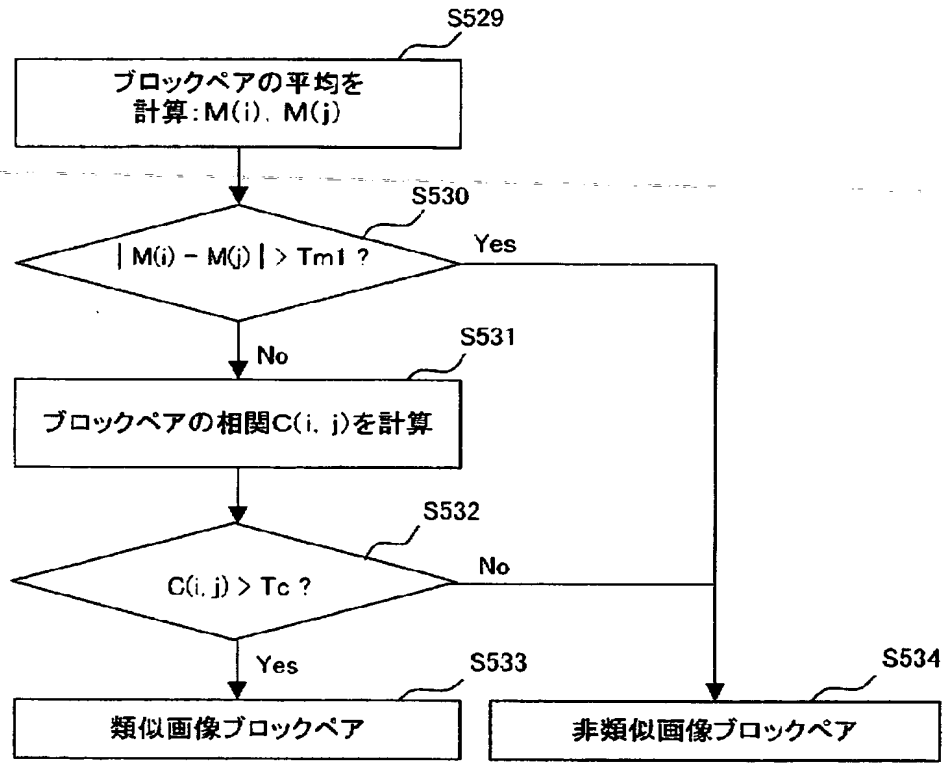
【図１２】

画像ブロック有効化部の動作フローチャート



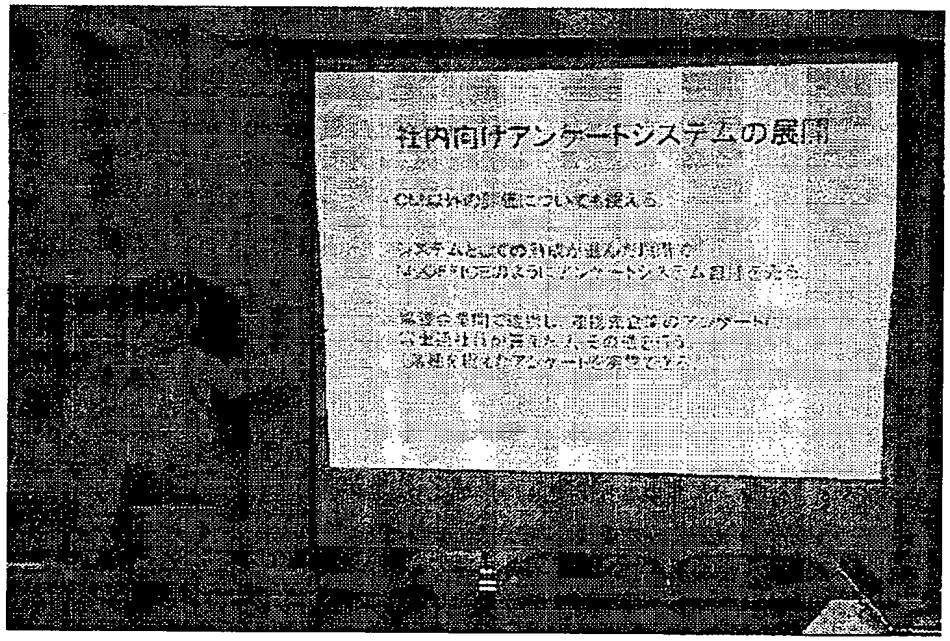
【図１３】

画像ブロック類似度計測部の動作フローチャート



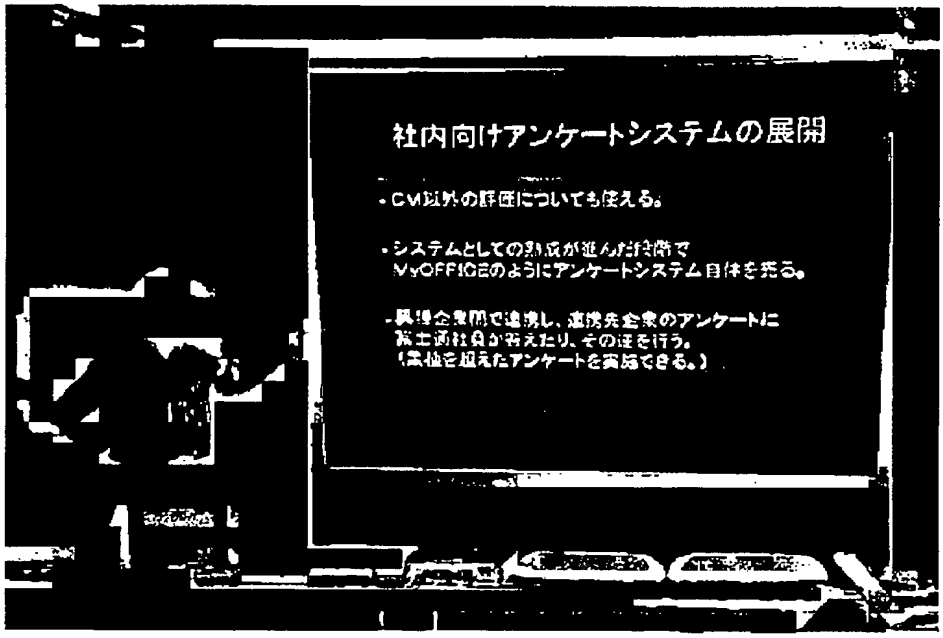
【図１４】

テキストフレーム検出検証のための
元のビデオフレームを示す図



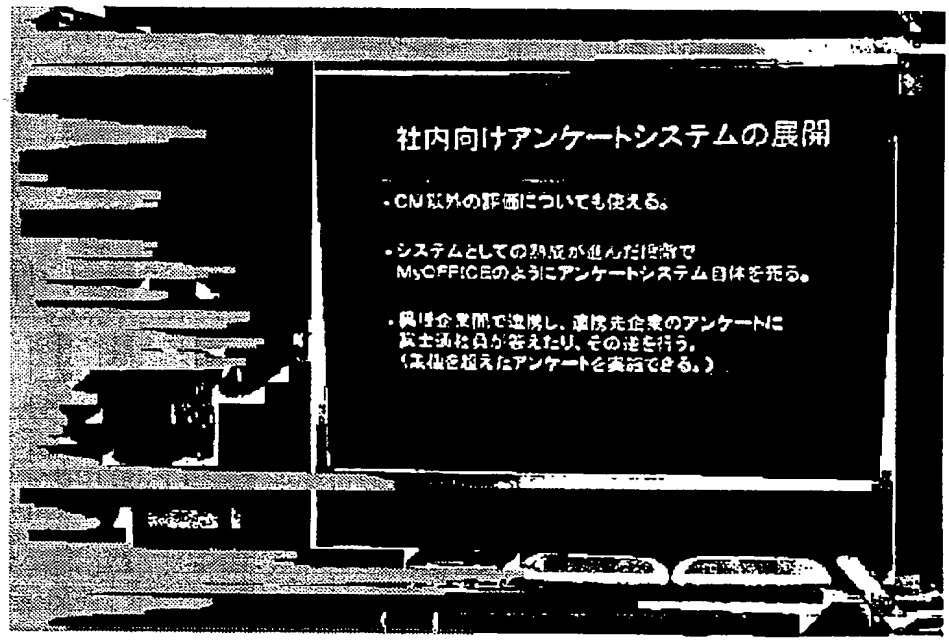
【図１５】

高速簡易画像二値化の結果の
第1の二値画像を示す図



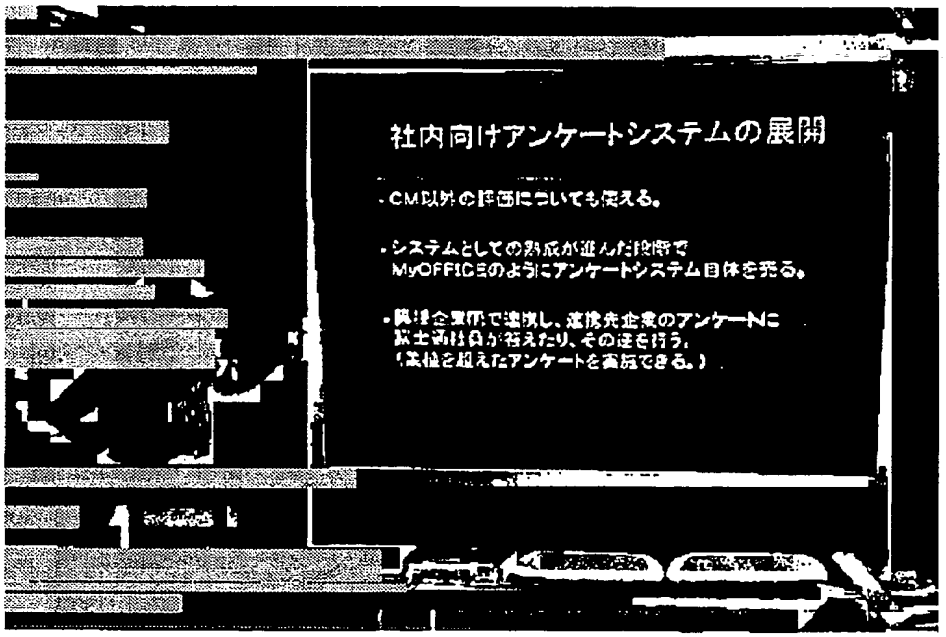
【図１６】

横射影の結果を示す図



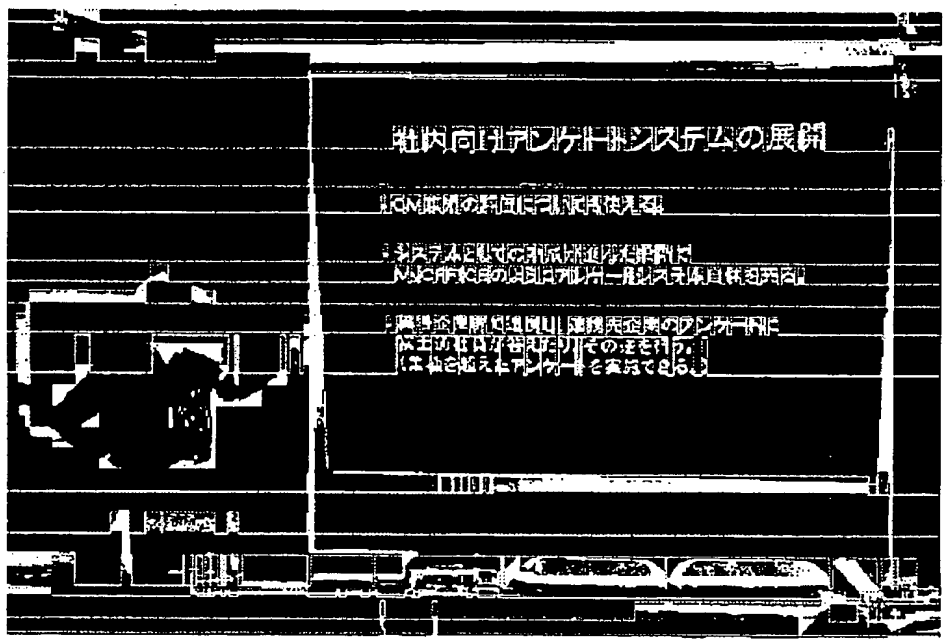
【図１７】

射影正則化の結果を示す図



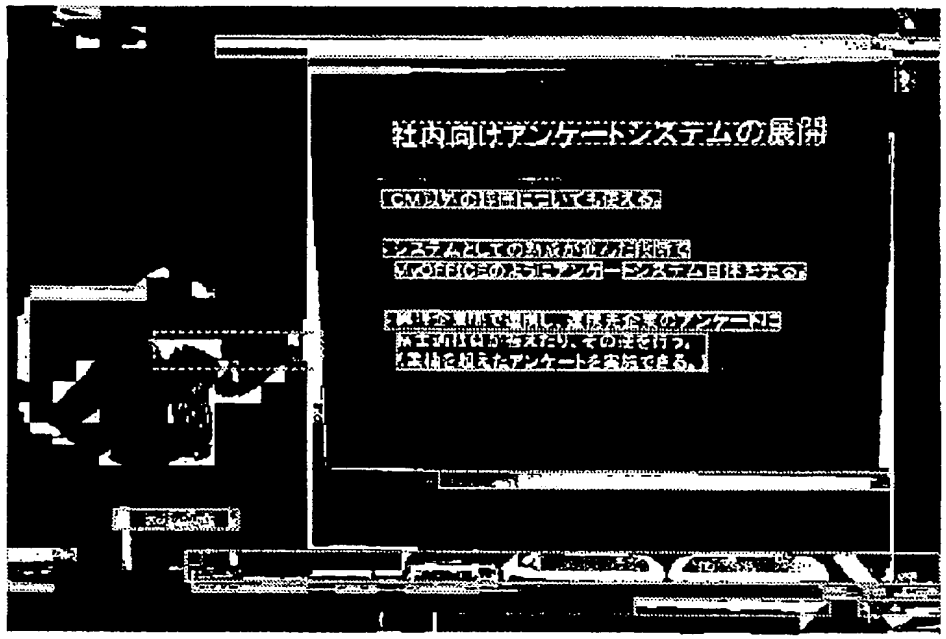
【図１８】

候補テキストライン毎の縦二値射影の結果を示す図



【図１９】

テキストライン領域決定の結果を示す図



【図２０】

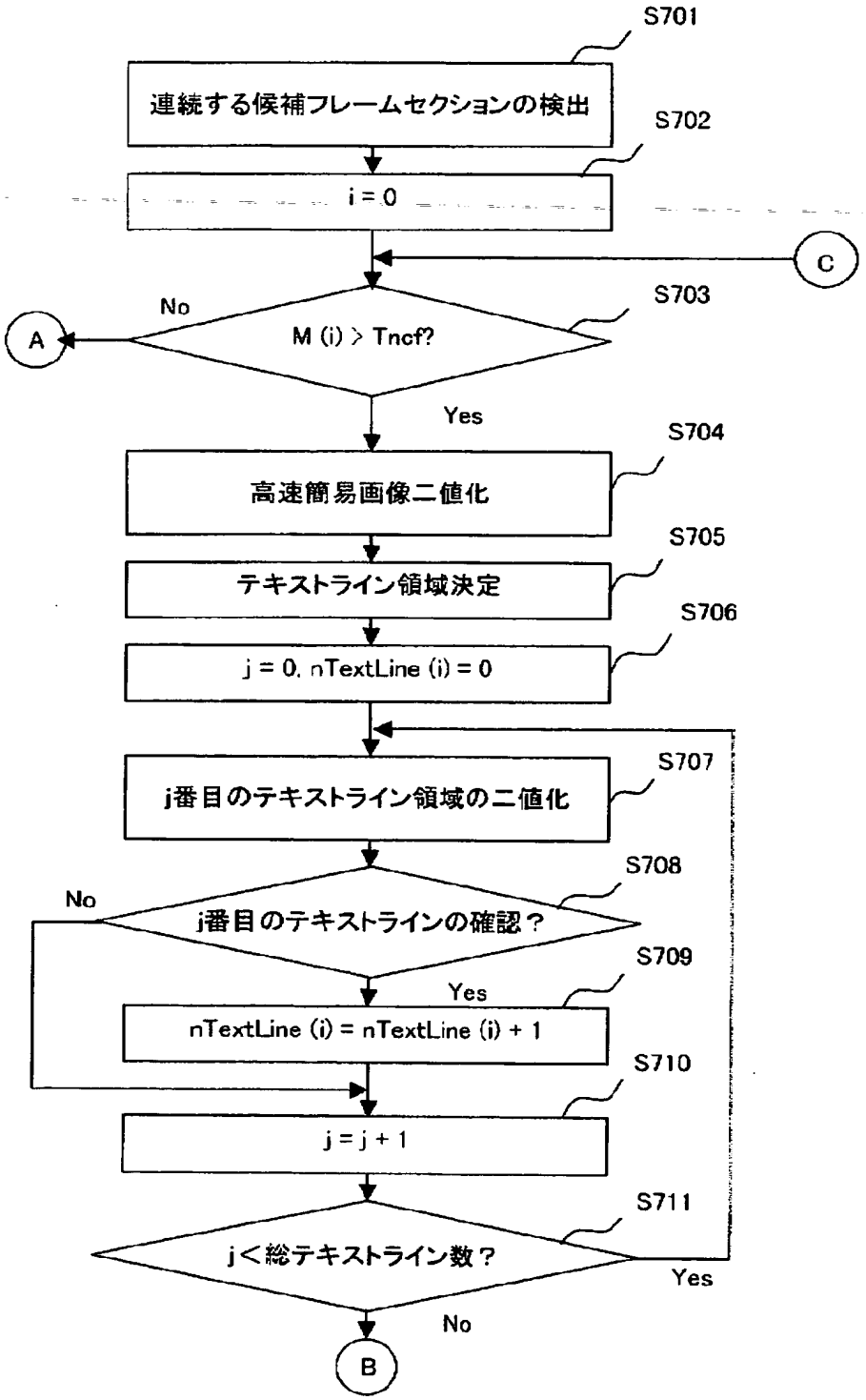
2つの候補テキストラインに
対する2組の二値画像を示す図

【図２１】

検出されたテキストライン領域を示す図

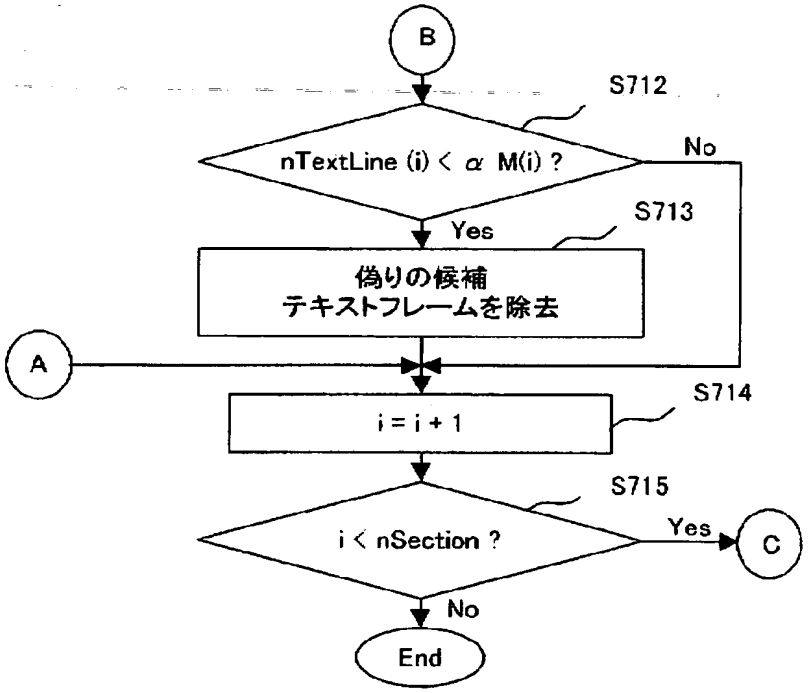
【図２２】

テキストフレーム検出検証部の
動作フローチャート(その1)



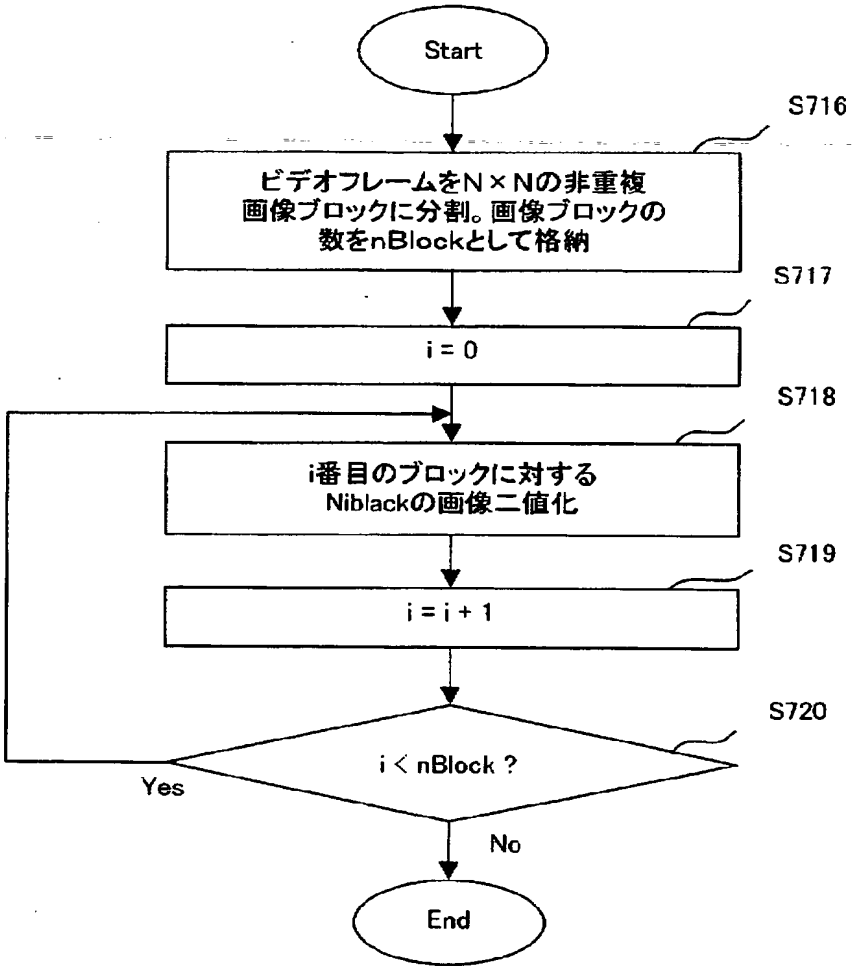
【図２３】

テキストフレーム検出検証部の
動作フローチャート(その2)



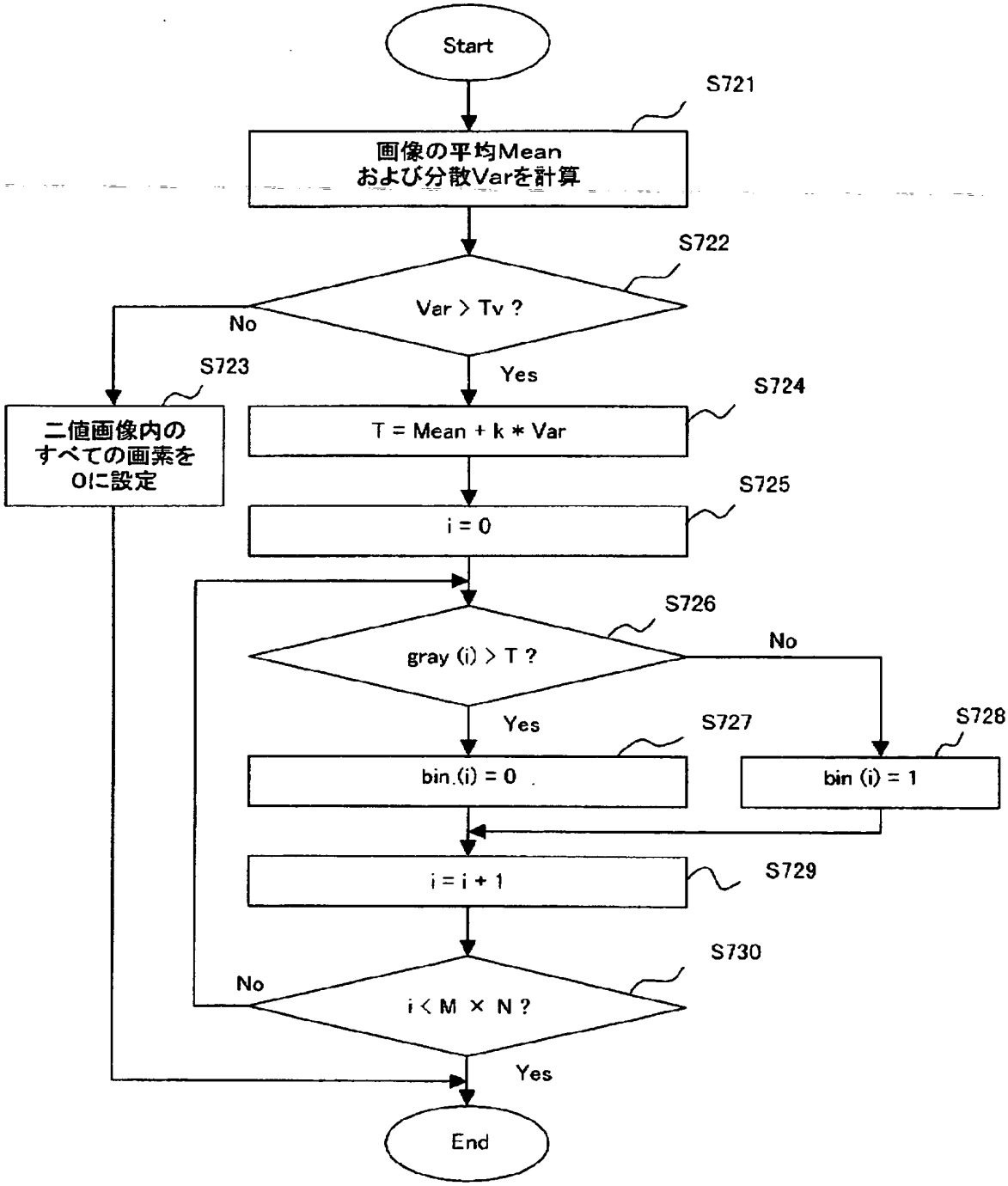
【図２４】

高速簡易二値化部の動作フローチャート



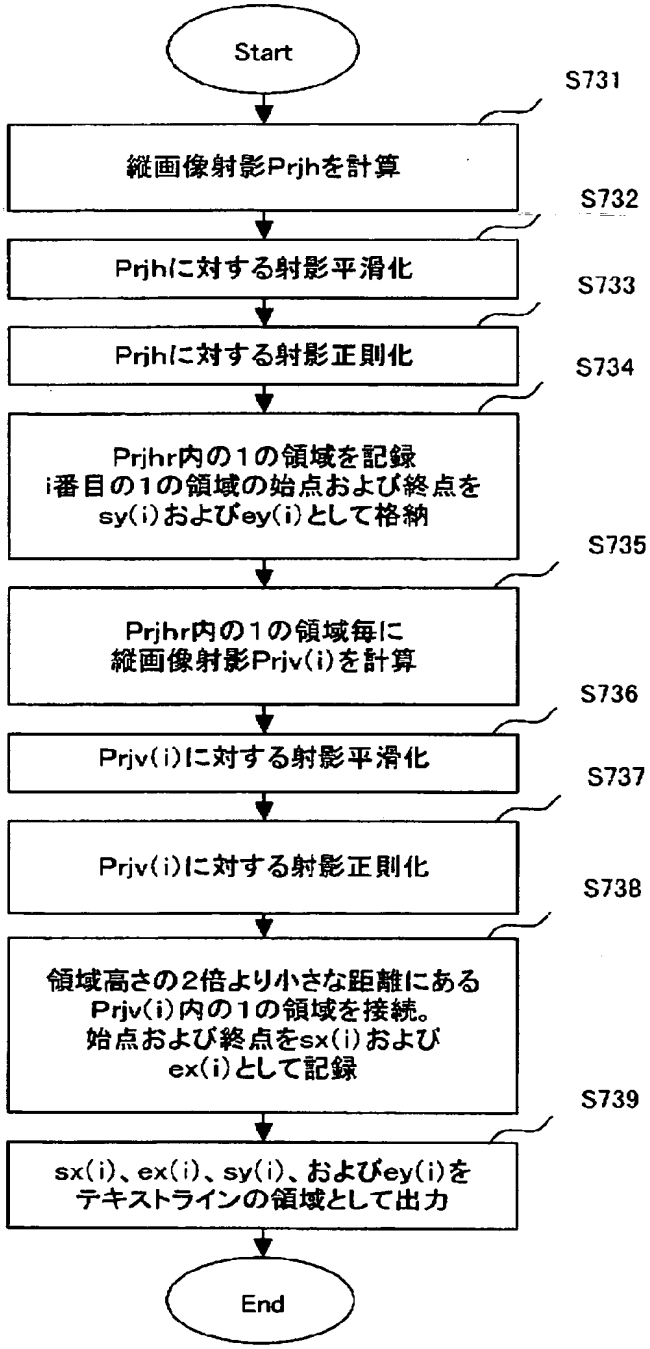
【図２５】

Niblackの画像二値化法のフローチャート



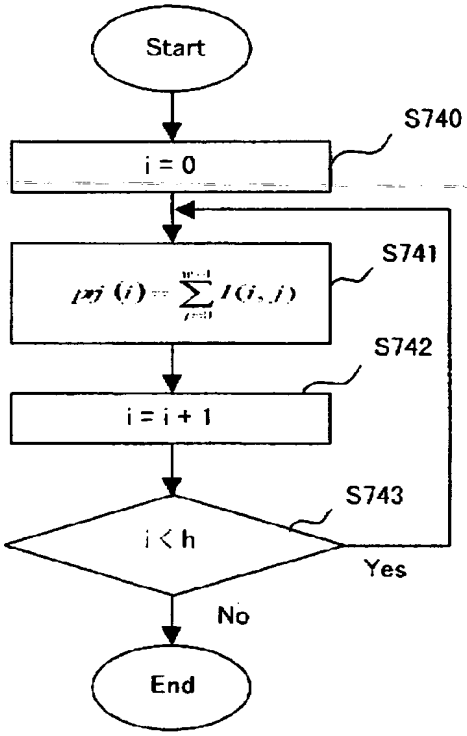
【図２６】

テキストライン領域決定部の動作フローチャート



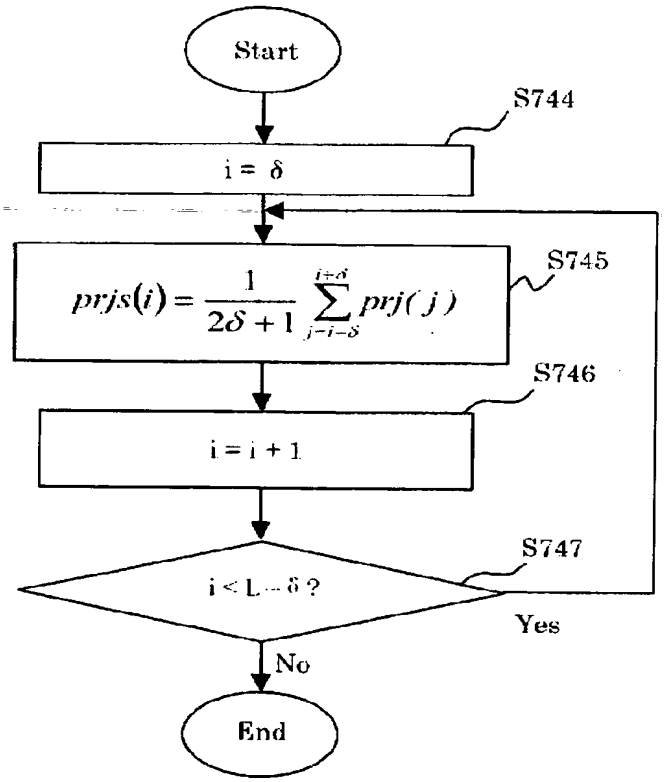
【図２７】

横画像射影のフローチャート



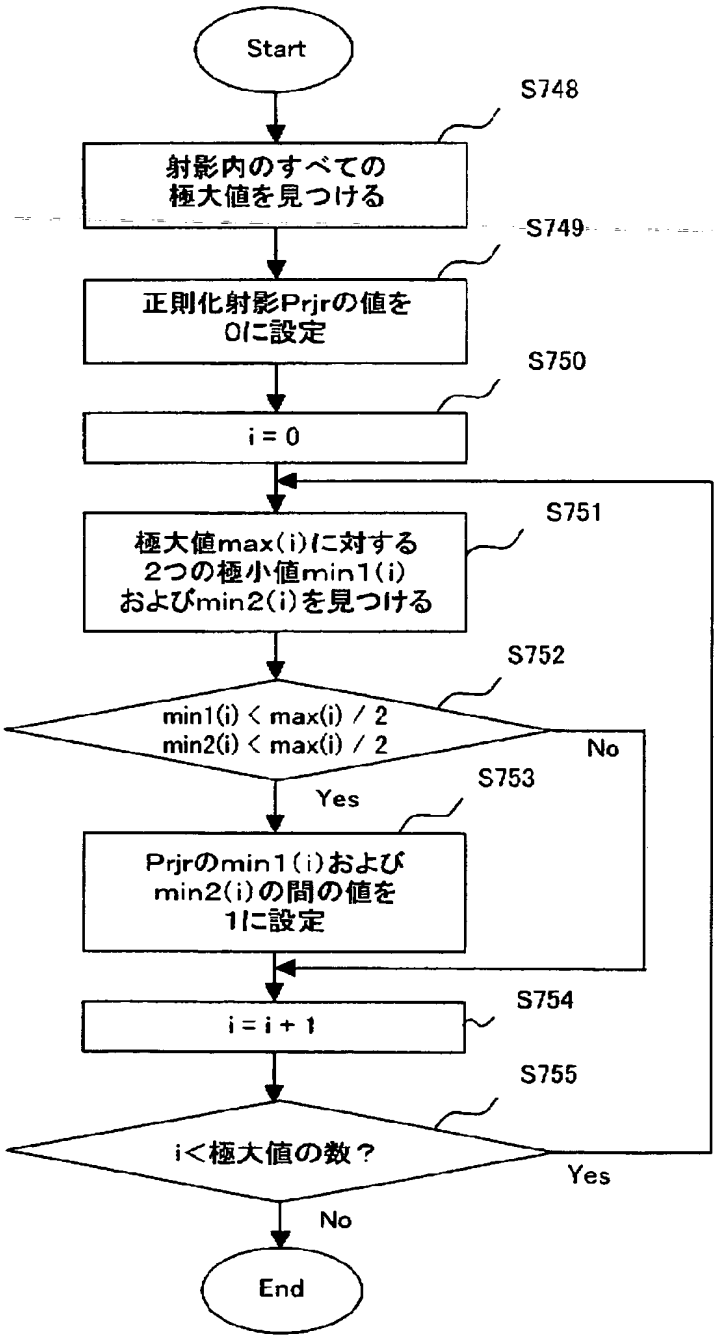
【図２８】

射影平滑化のフローチャート



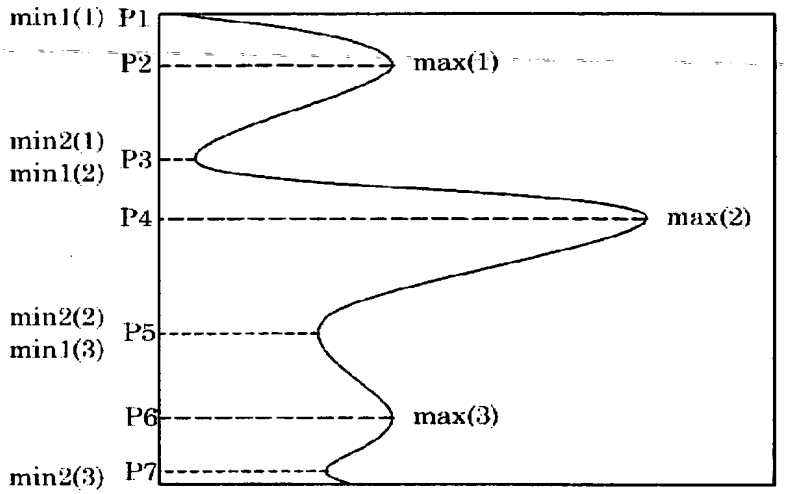
【図２９】

射影正則化のフローチャート



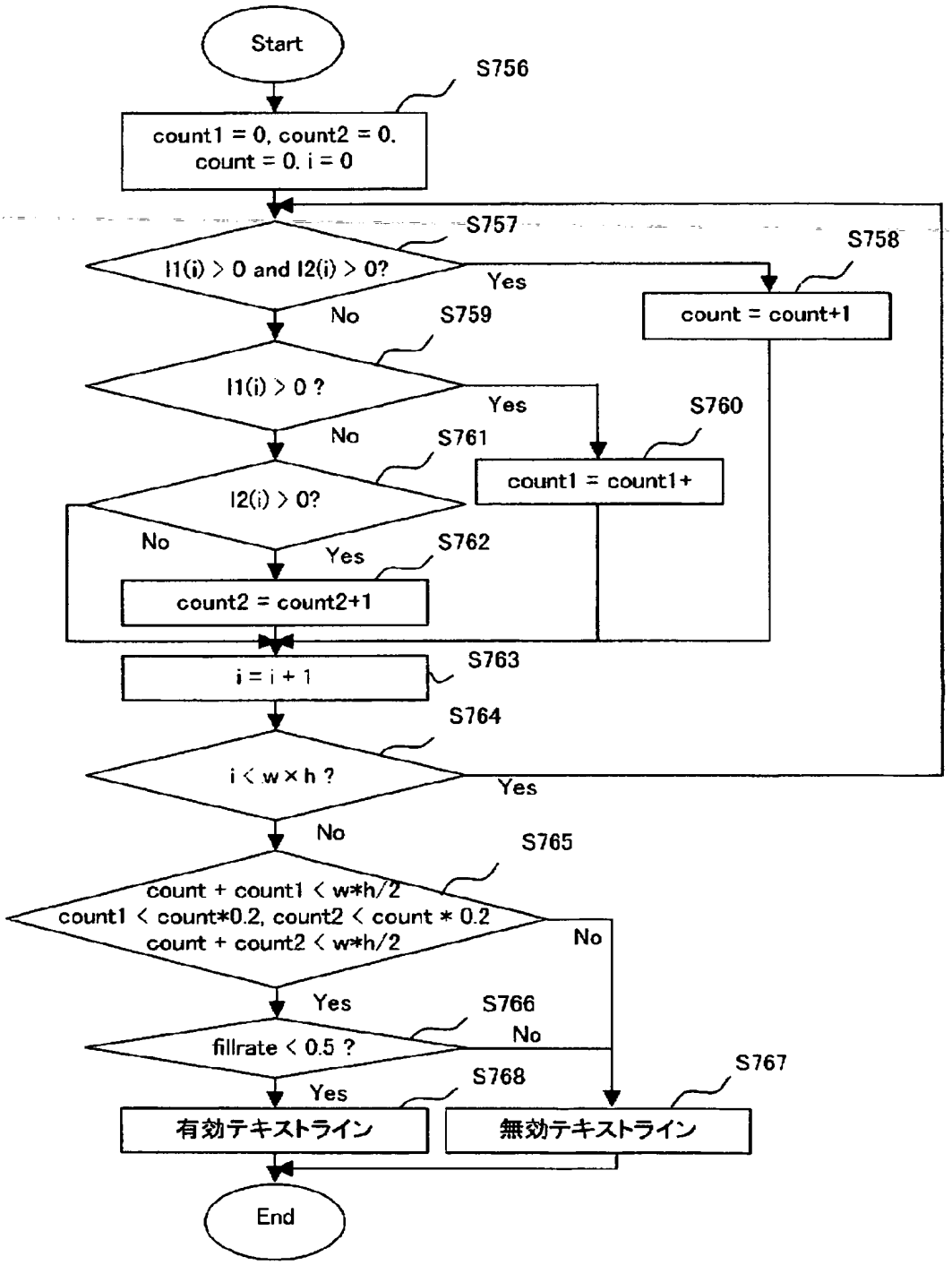
【図３０】

射影のmaxおよびminの例を示す図



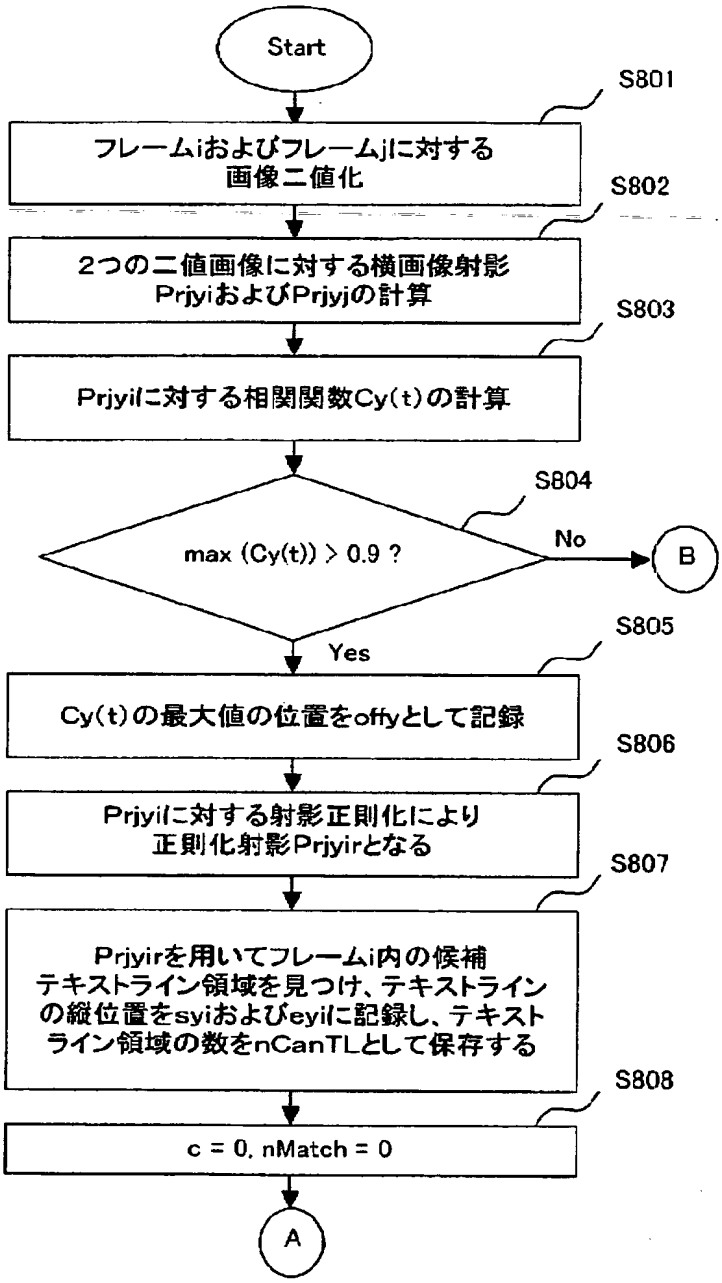
【図３１】

テキストライン確認部の動作フローチャート



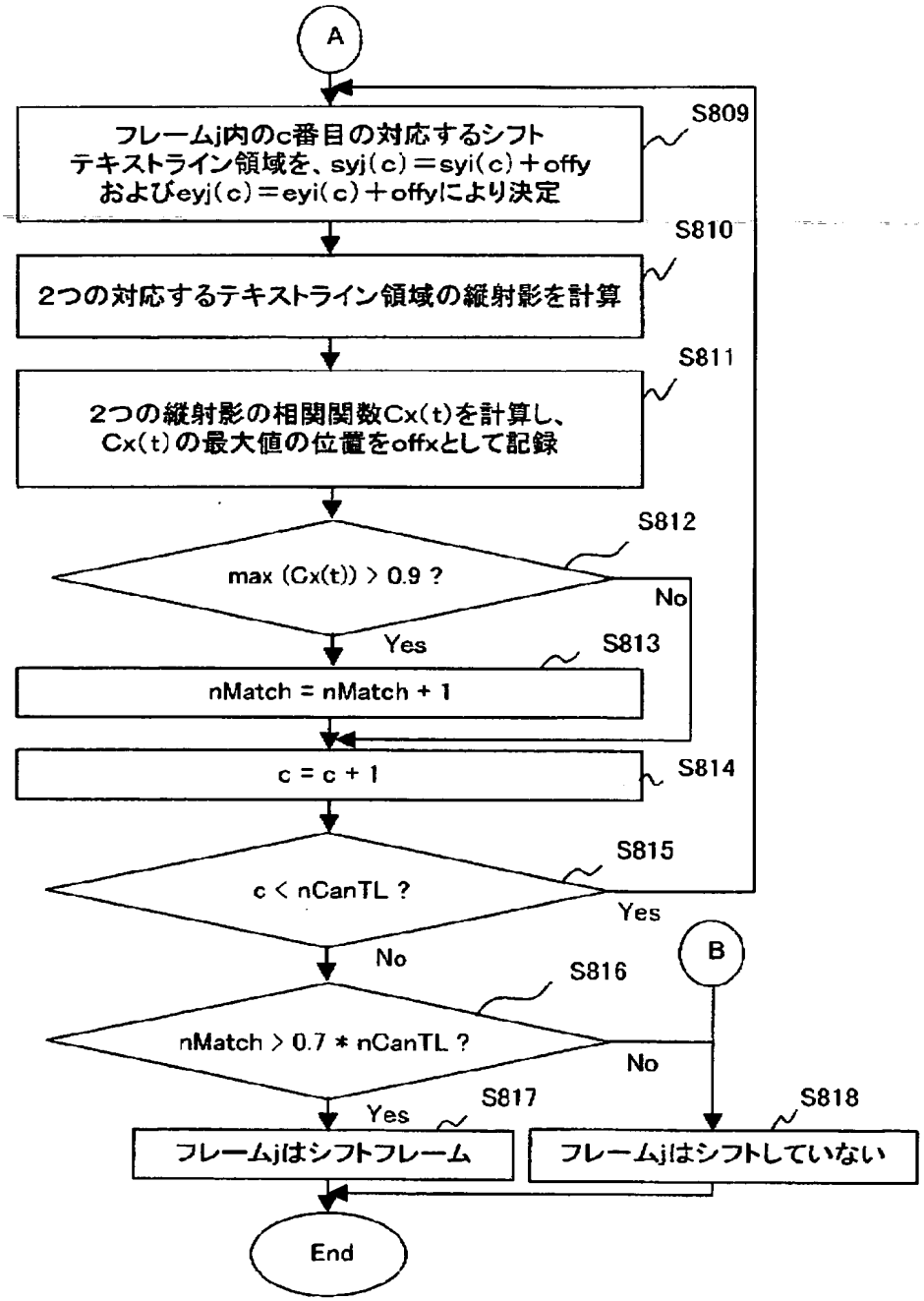
【図３２】

画像シフト検出部の動作フローチャート（その１）



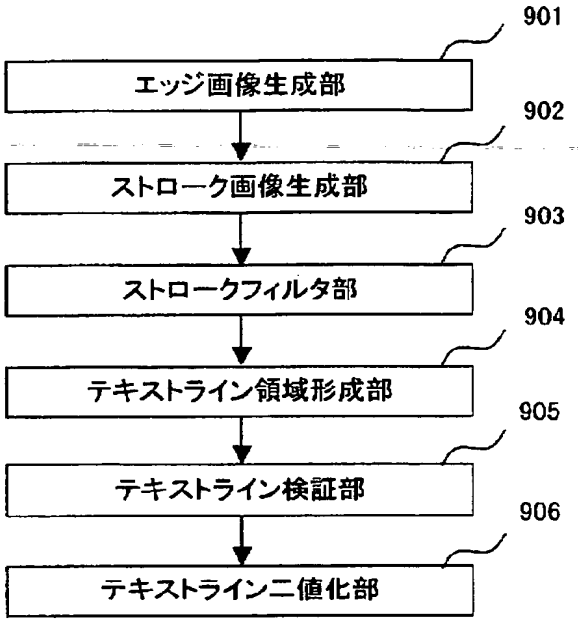
【図３３】

画像シフト検出部の動作フローチャート（その2）



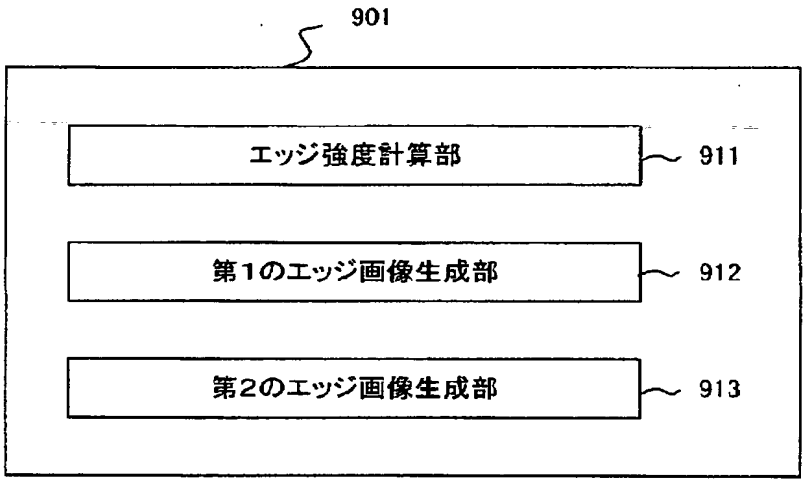
【図３４】

本発明のテキスト抽出装置の構成を示す図



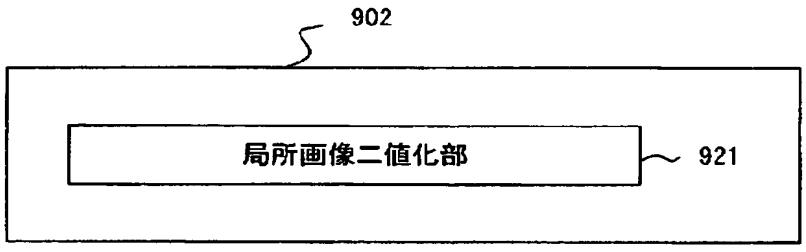
【図３５】

エッジ画像生成部の構成を示す図



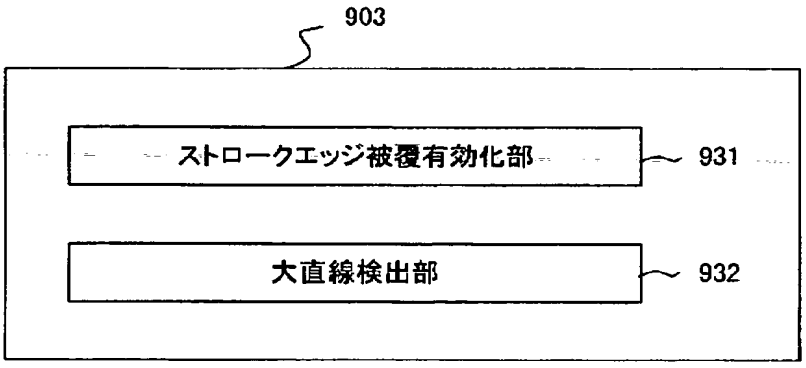
【図３６】

ストローク画像生成部の構成を示す図



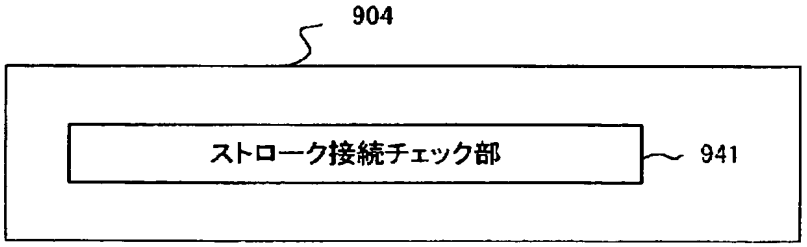
【図３７】

ストロークフィルタ部の構成を示す図



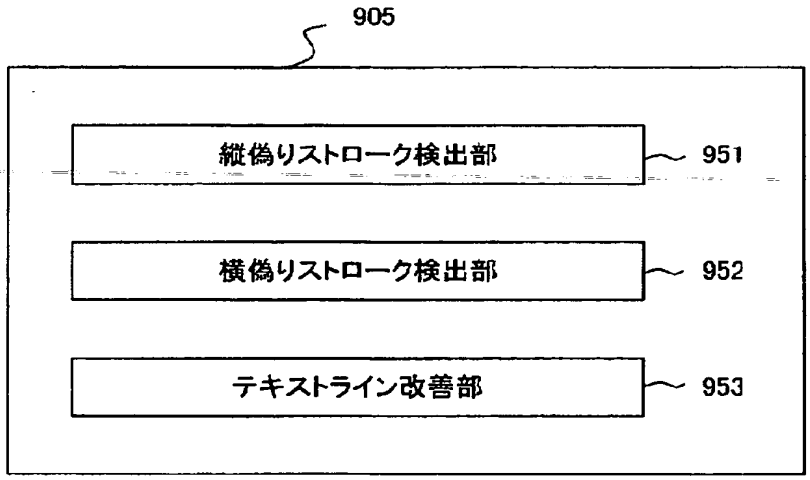
【図３８】

テキストライン領域形成部の構成を示す図



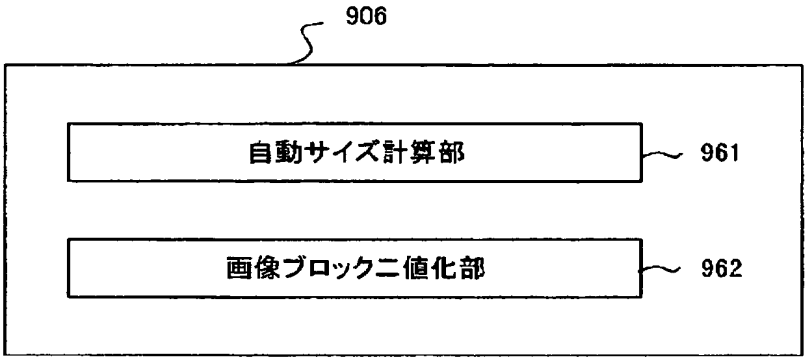
【図３９】

テキストライン検証部の構成を示す図



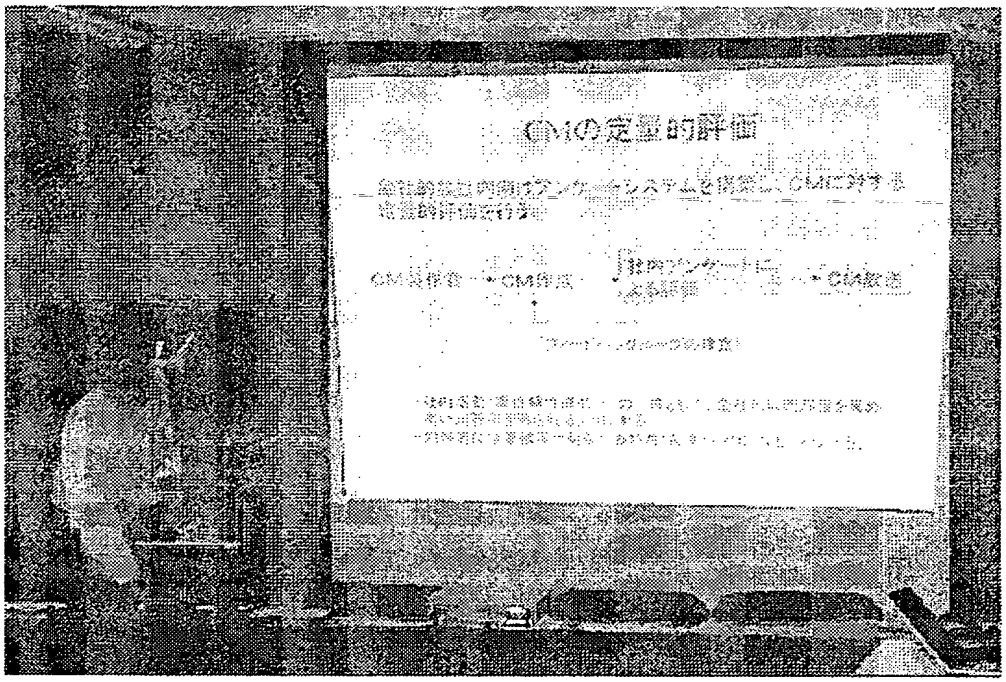
【図４０】

テキストライン二値化部の構成を示す図



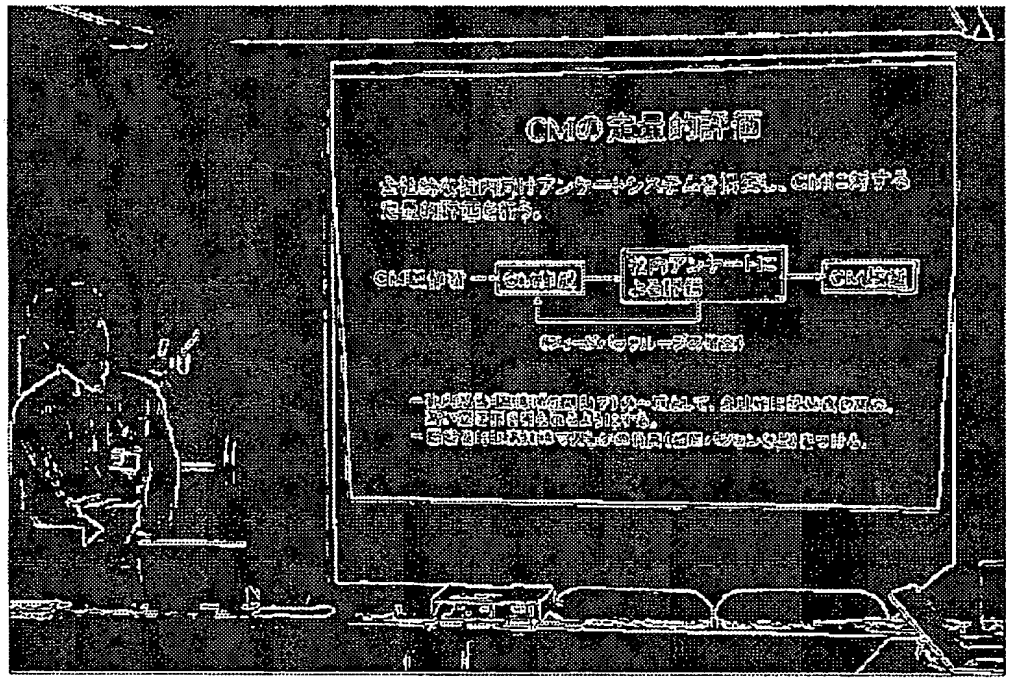
【図４１】

テキスト抽出のための元のビデオフレームを示す図



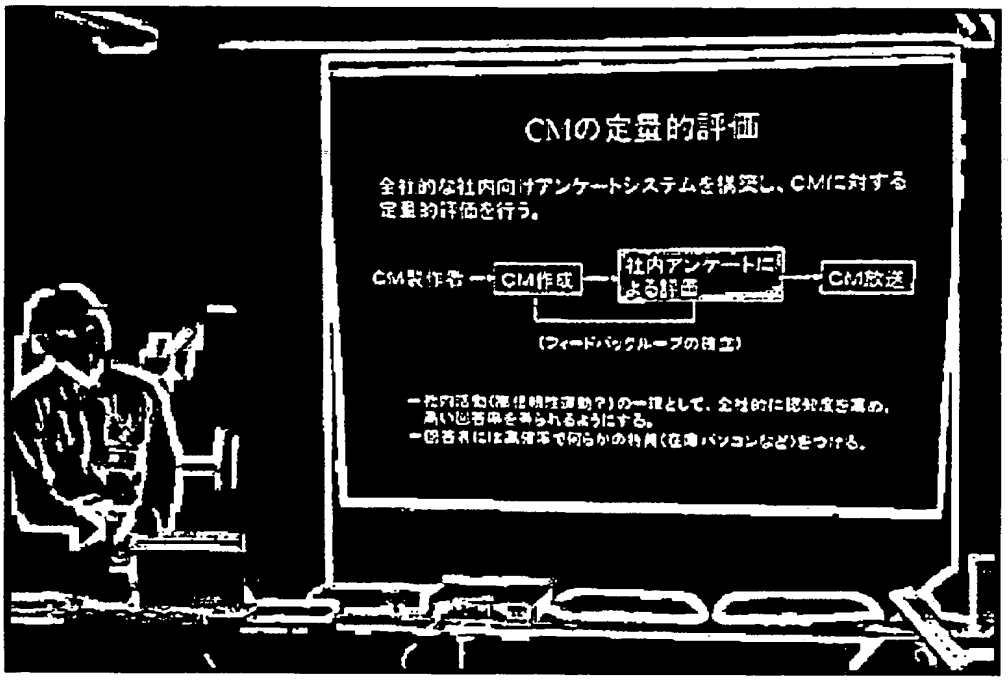
【図４２】

エッジ画像生成の結果を示す図



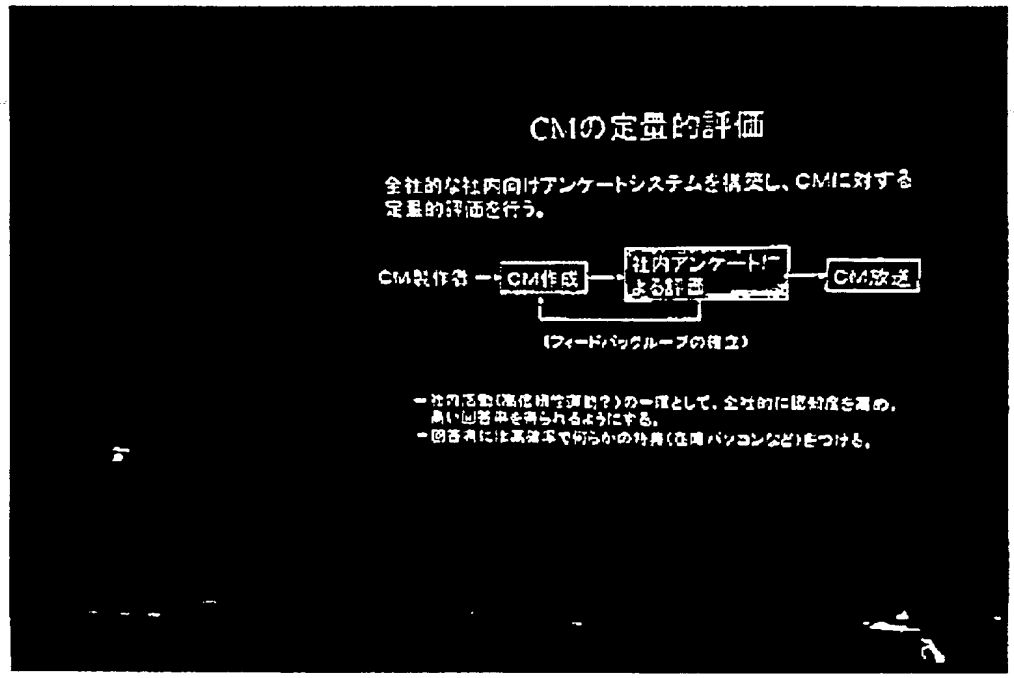
【図４３】

ストローク生成の結果を示す図



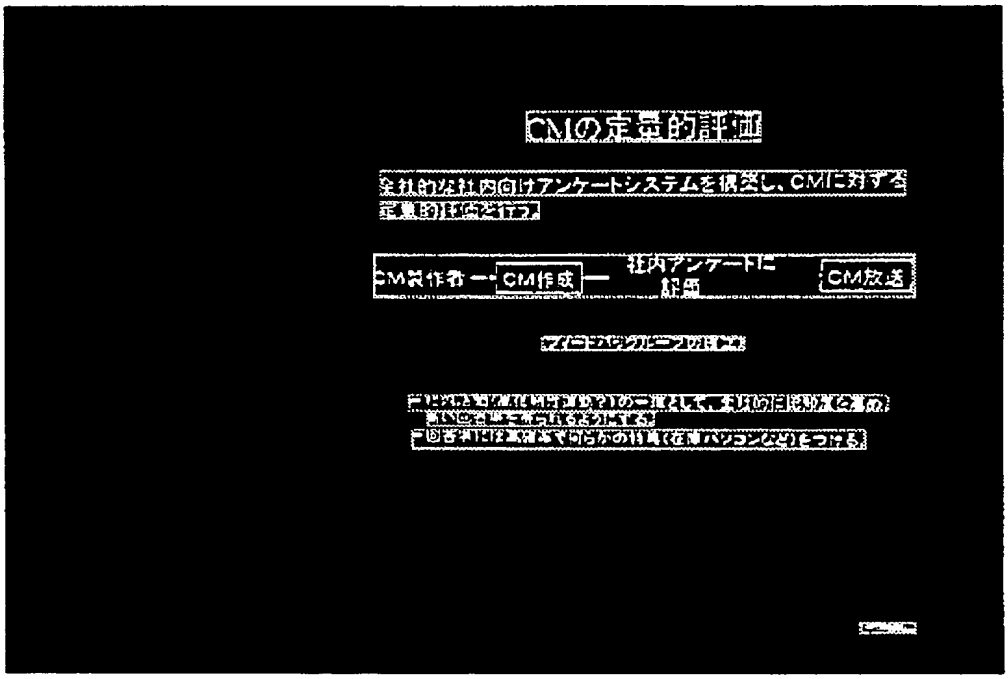
【図４４】

ストロークフィルタリングの結果を示す図



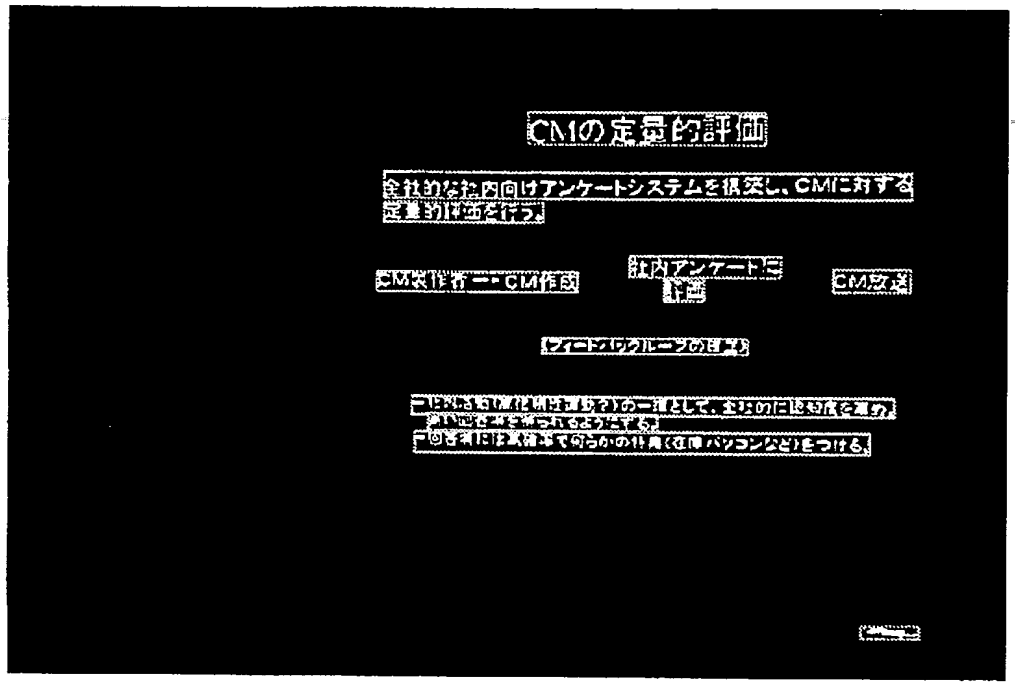
【図４５】

テキストライン領域形成の結果を示す図



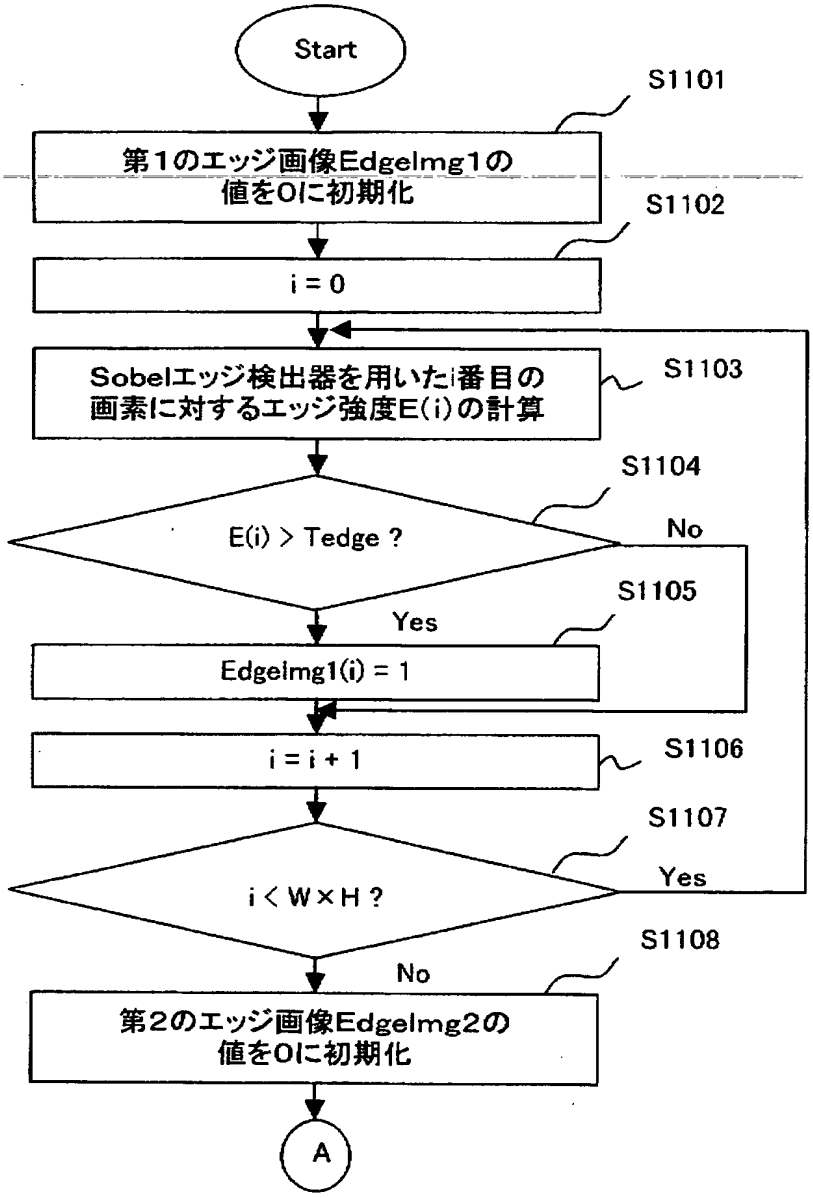
【図４６】

最終二値化テキストライン領域を示す図



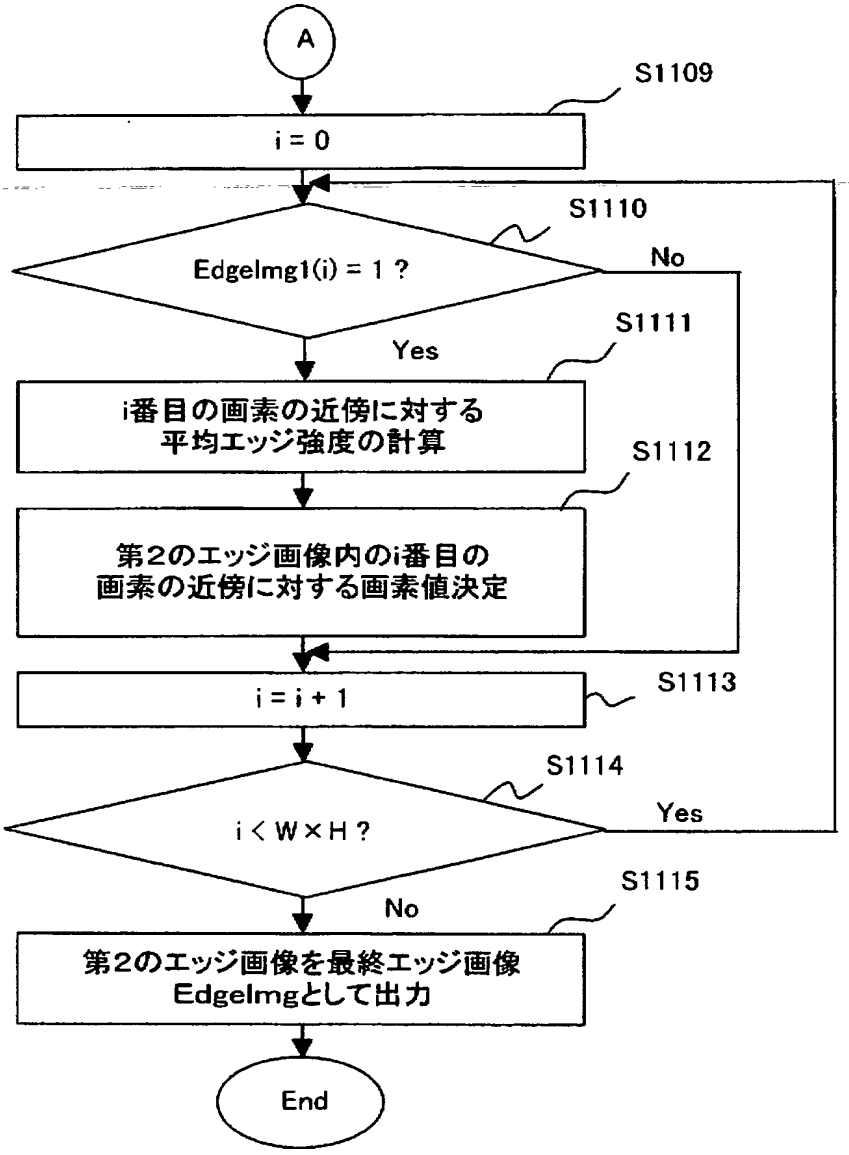
【図４７】

エッジ画像生成部の動作フローチャート（その１）



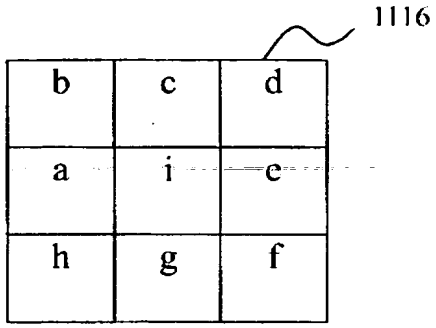
【図４８】

エッジ画像生成部の動作フローチャート（その２）



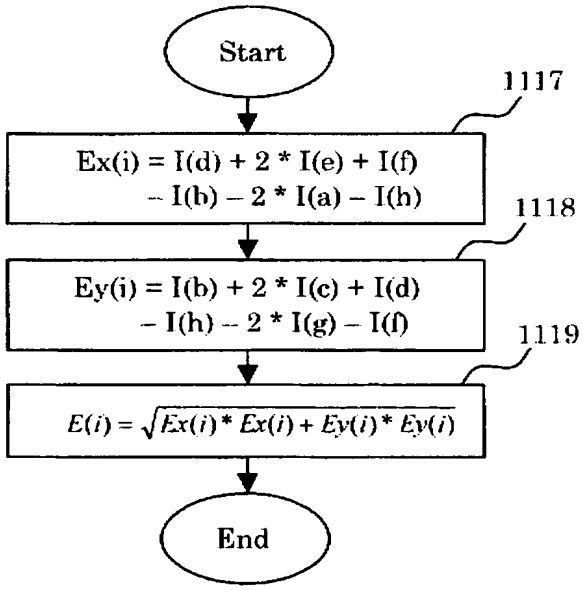
【図４９】

画素iの近傍の配置を示す図



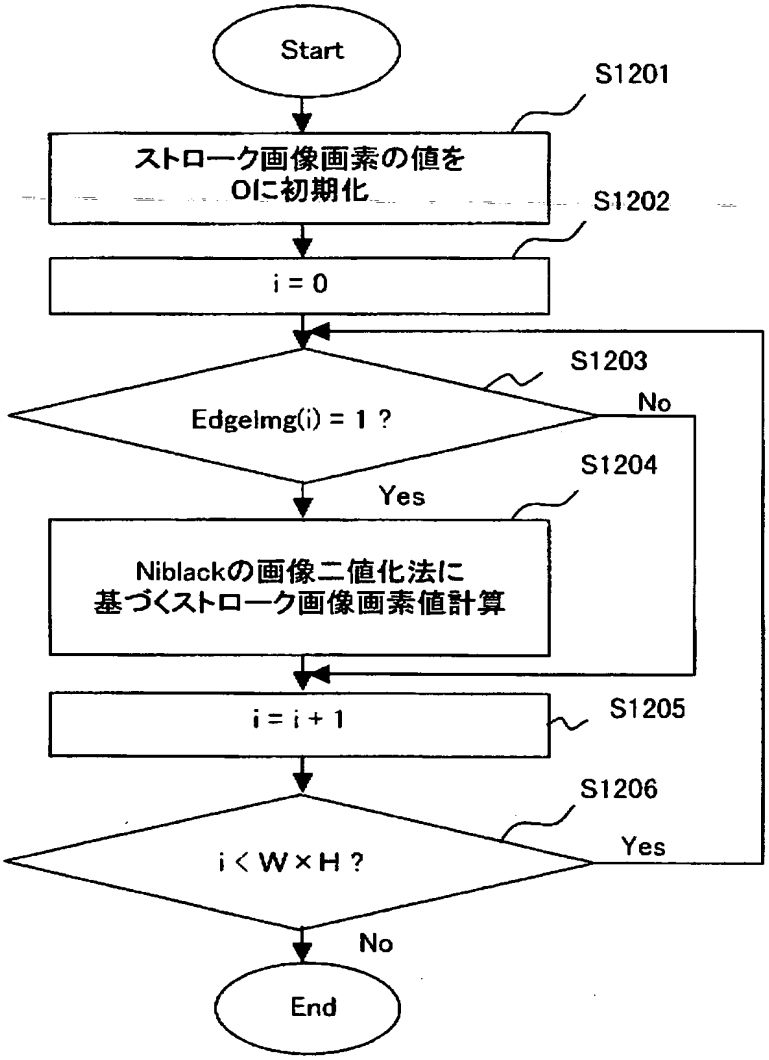
【図５０】

エッジ強度計算部の動作フローチャート



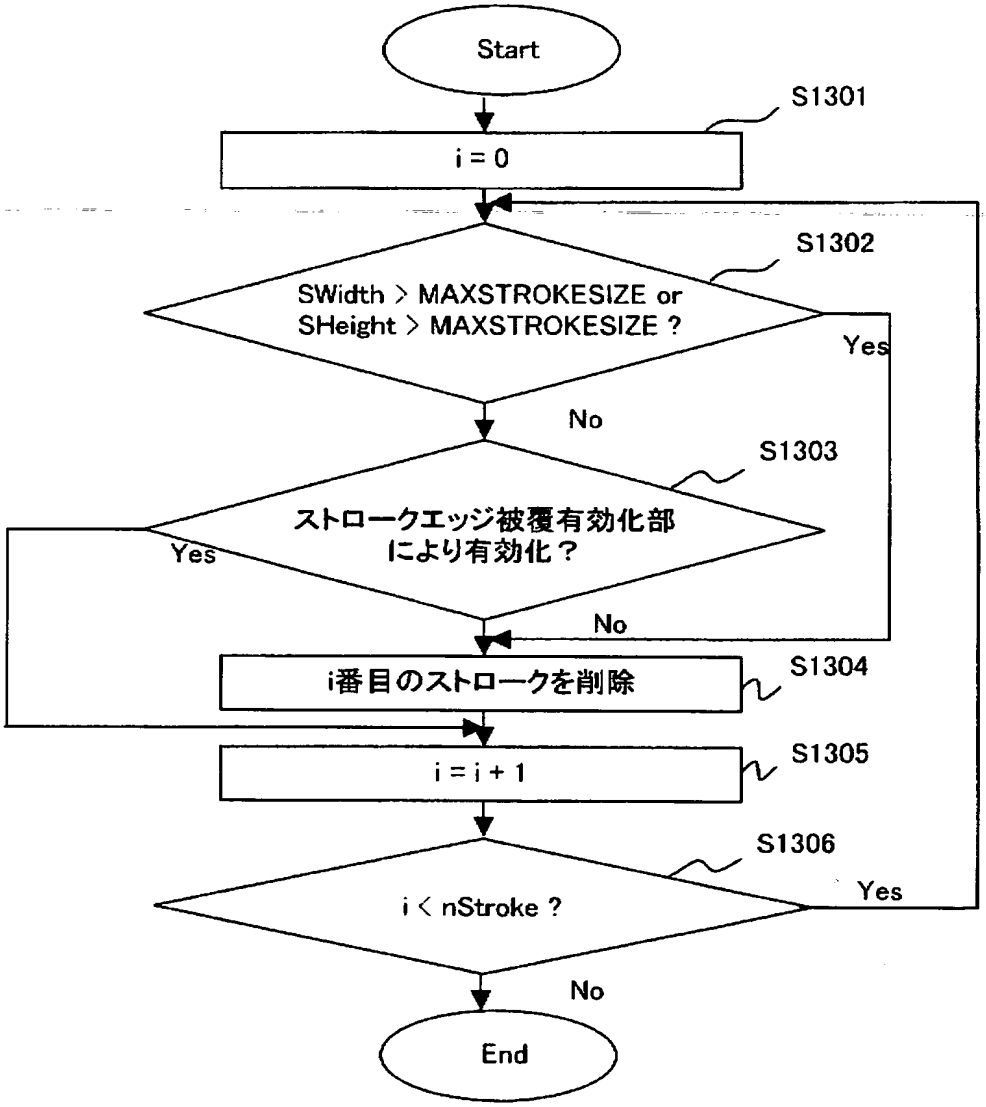
【図５１】

ストローク画像生成部の動作フローチャート



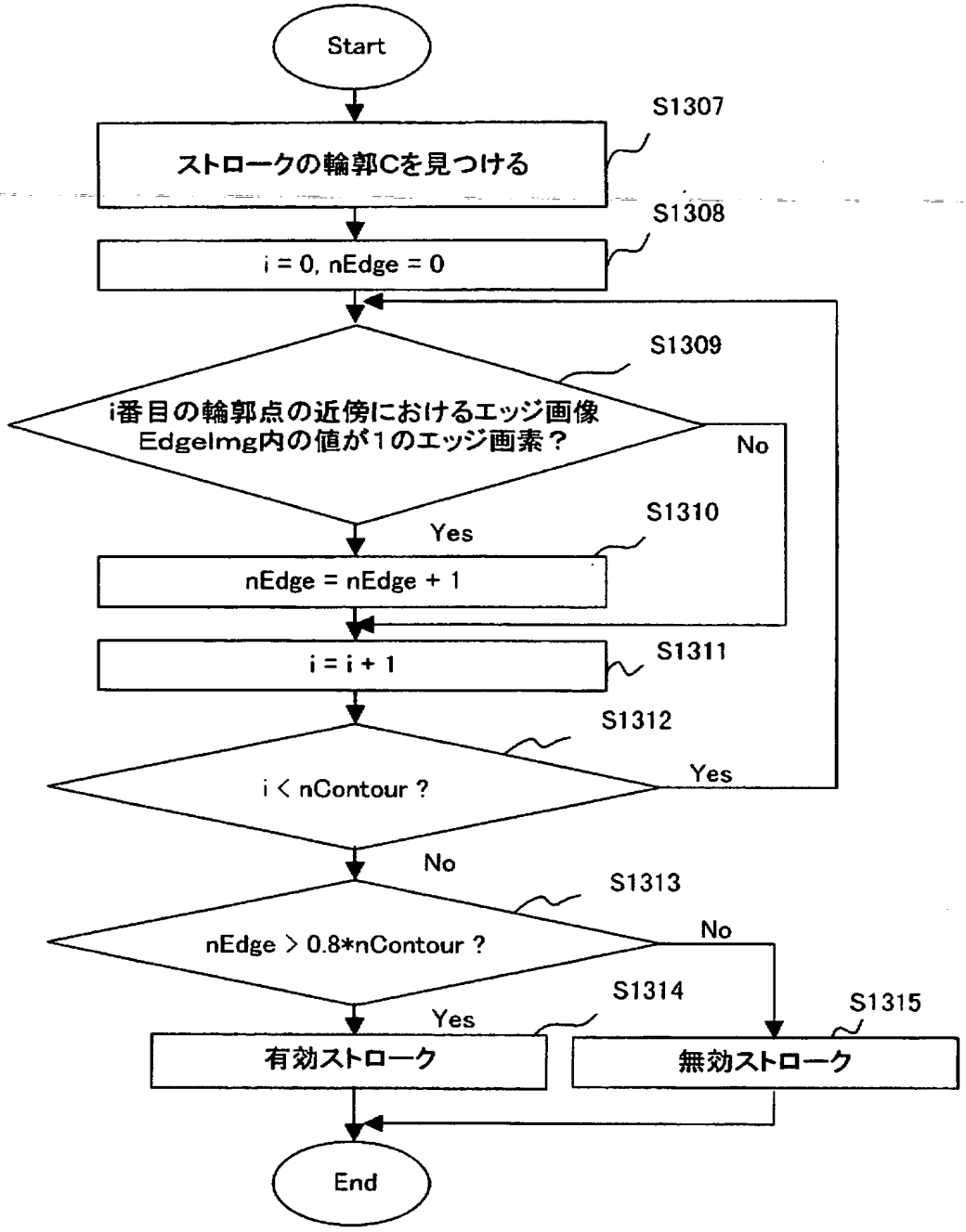
【図５２】

ストロークフィルタ部の動作フローチャート



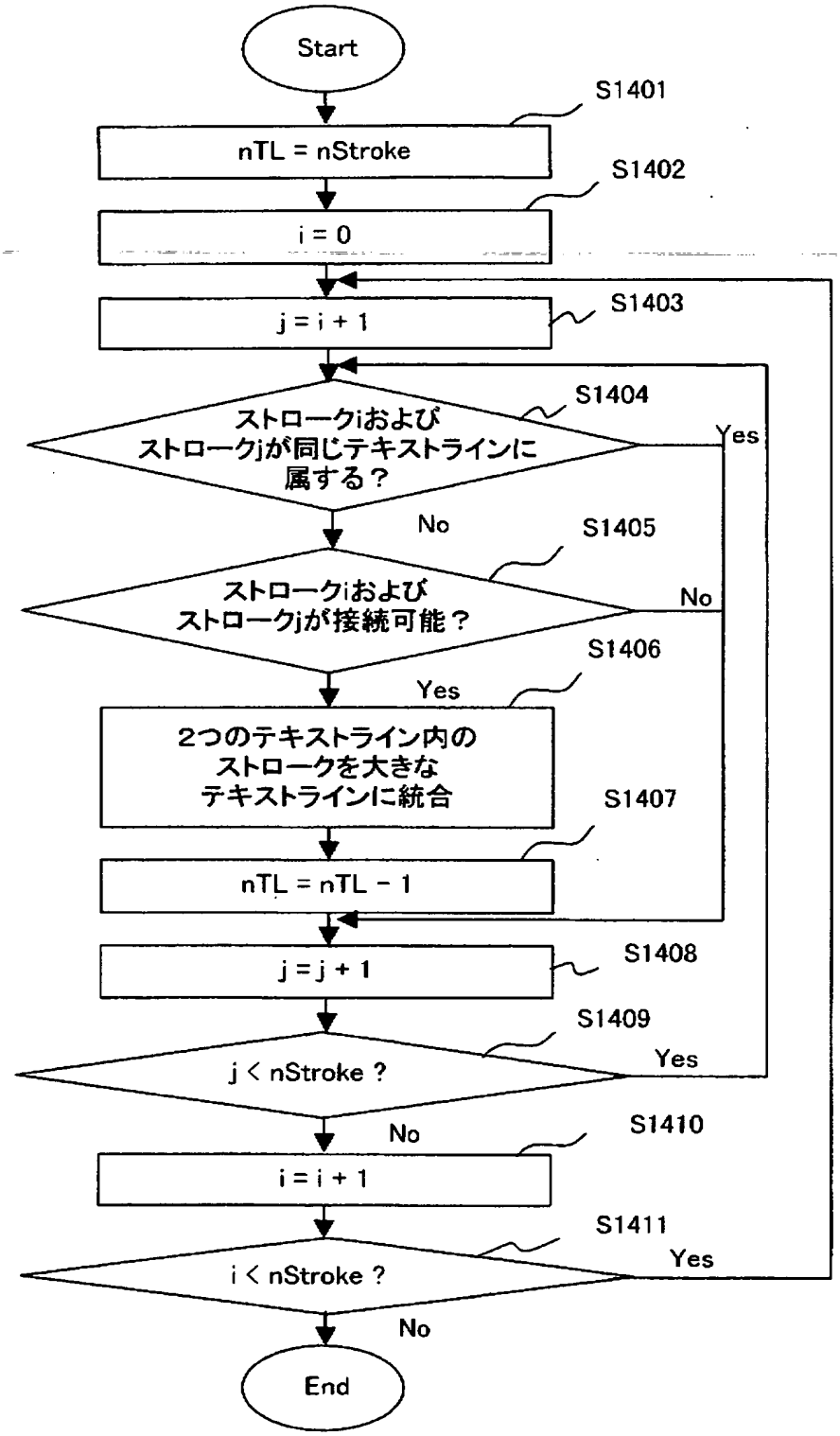
【図５３】

ストロークエッジ被覆有効化部の動作フローチャート



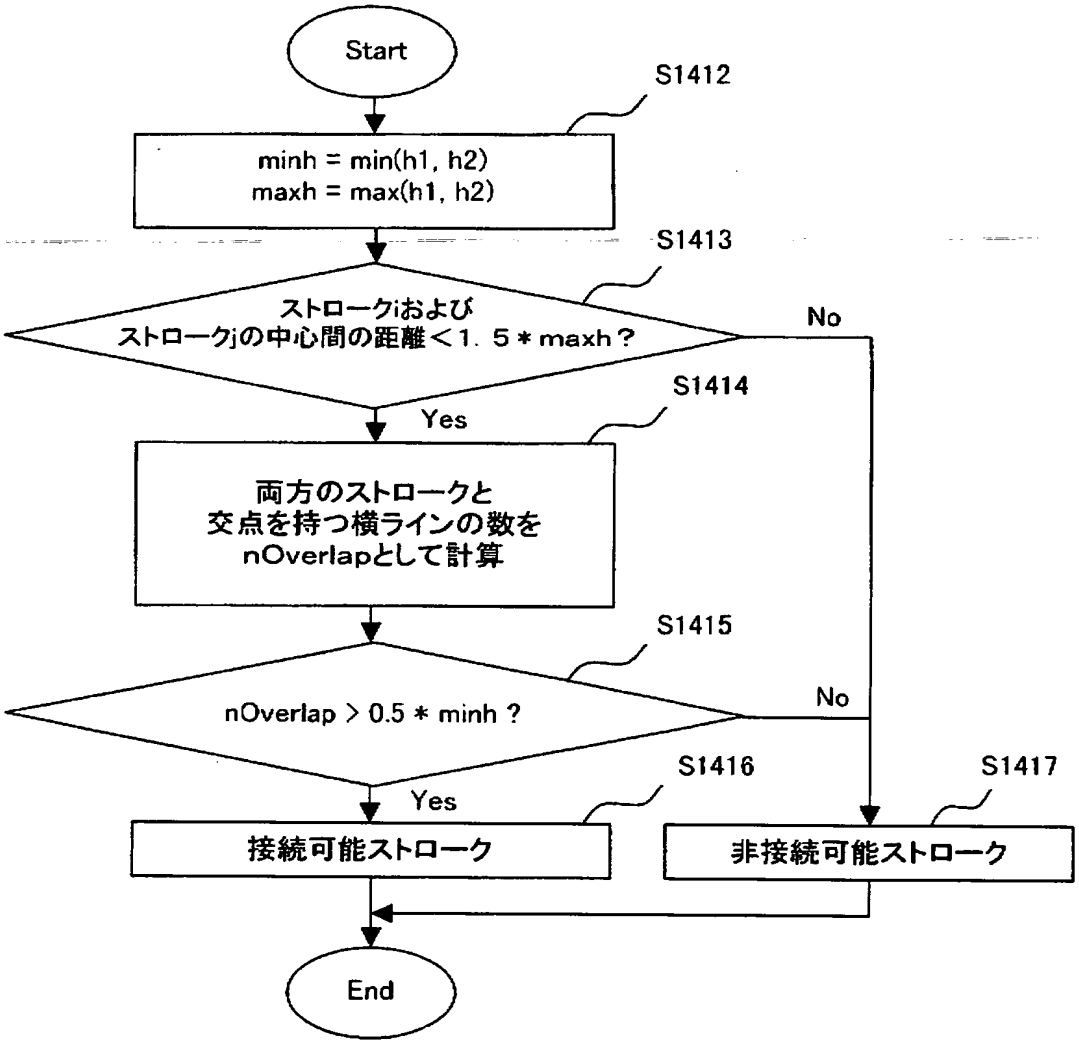
【図５４】

テキストライン領域形成部の動作フローチャート



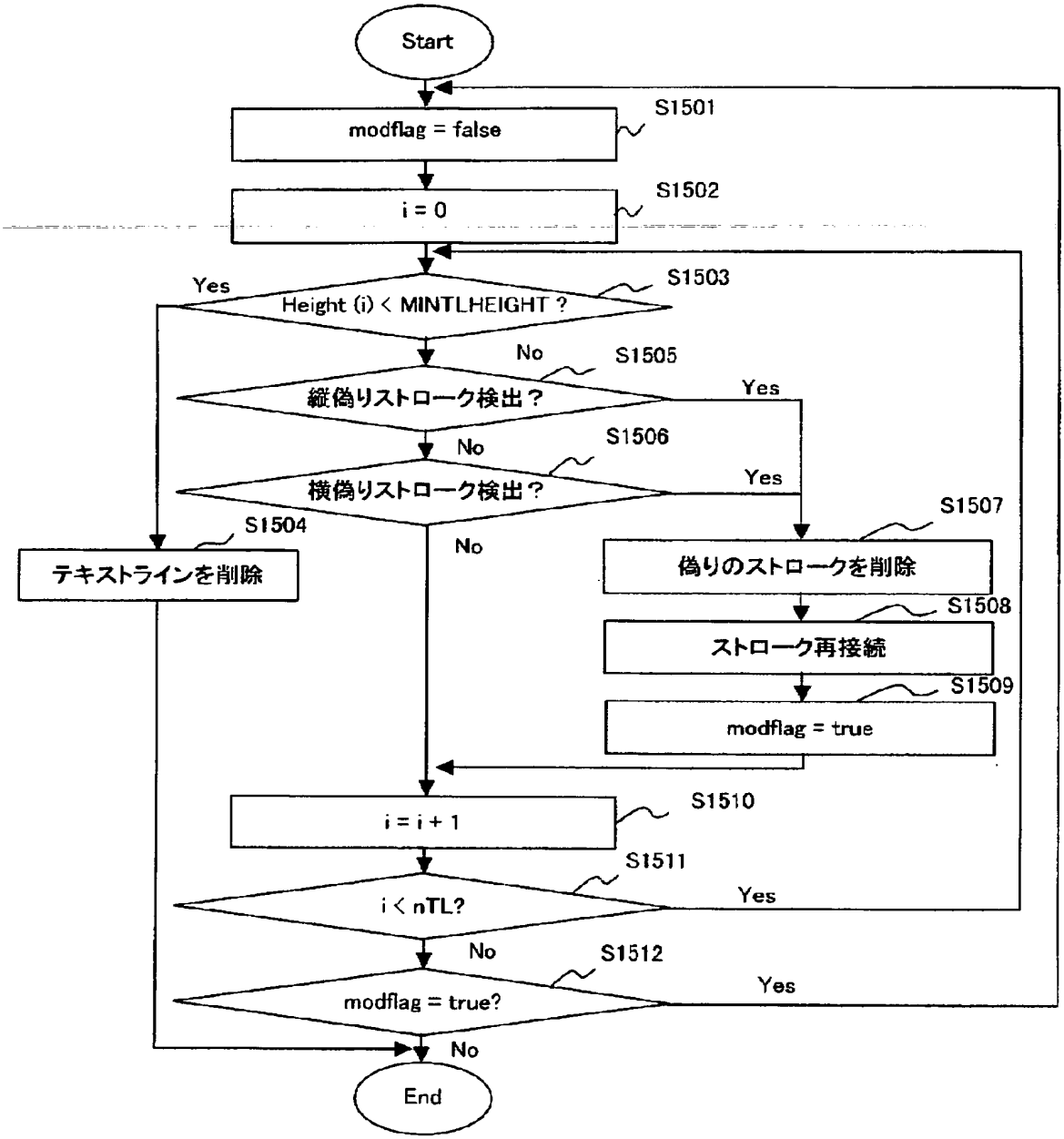
【図５５】

ストローク接続チェック部の動作フローチャート



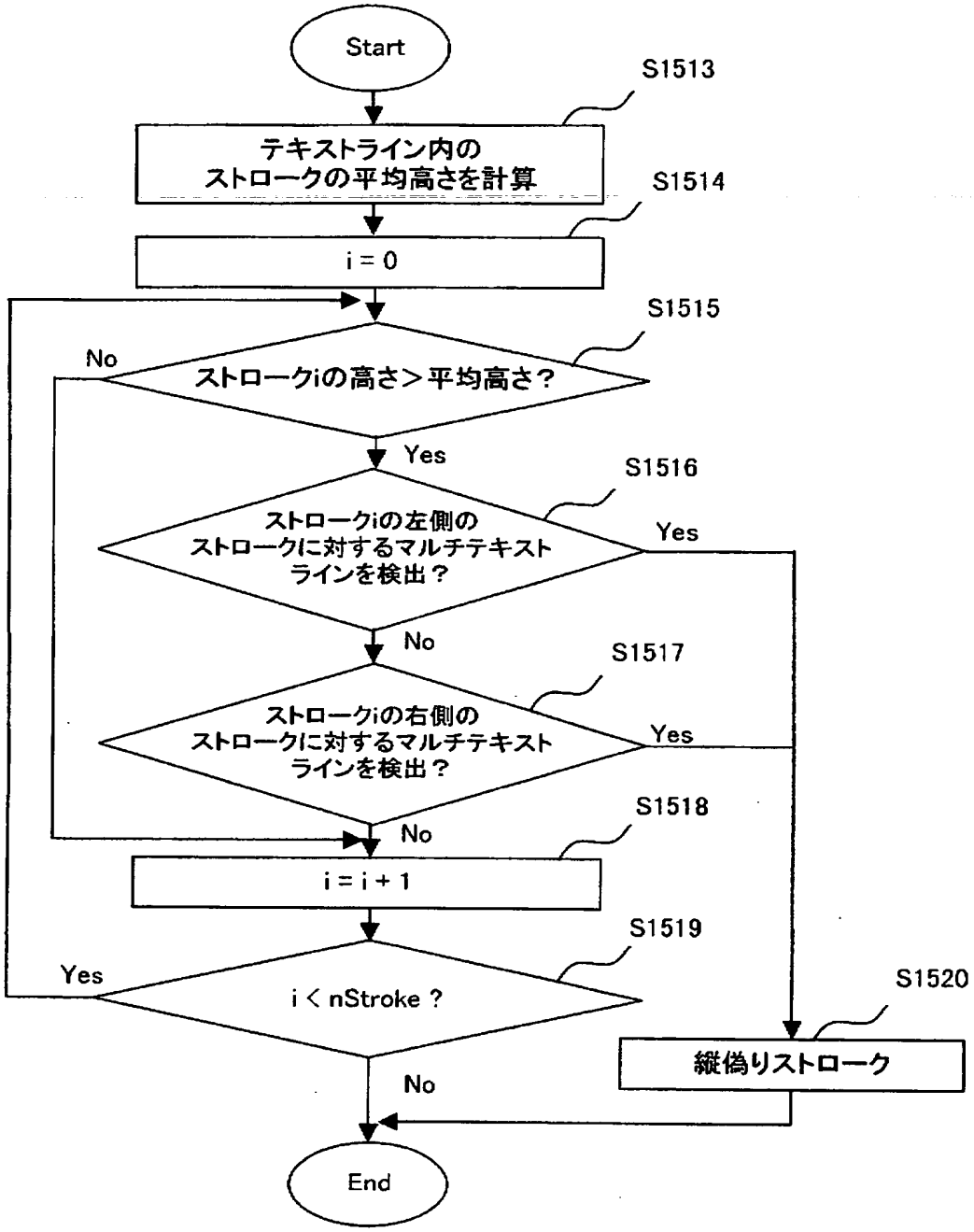
【図５６】

テキストライン検証部の動作フローチャート



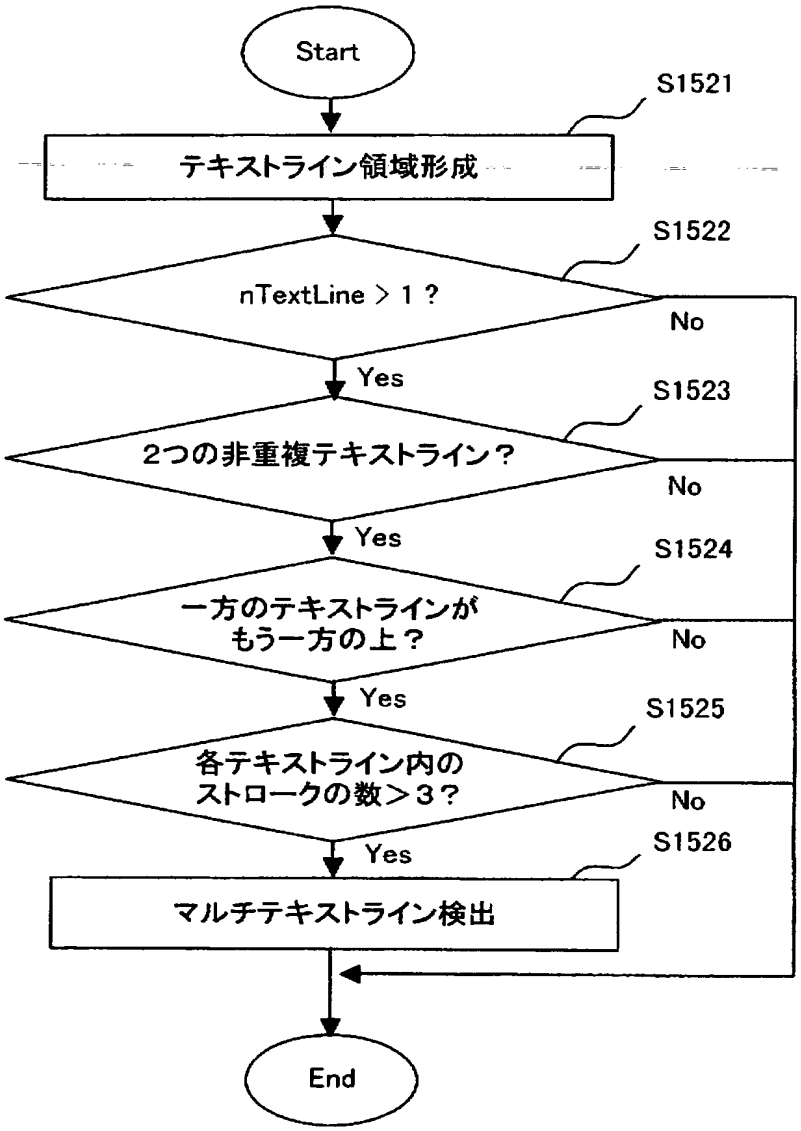
【図５７】

縦偽りストローク検出部の動作フローチャート



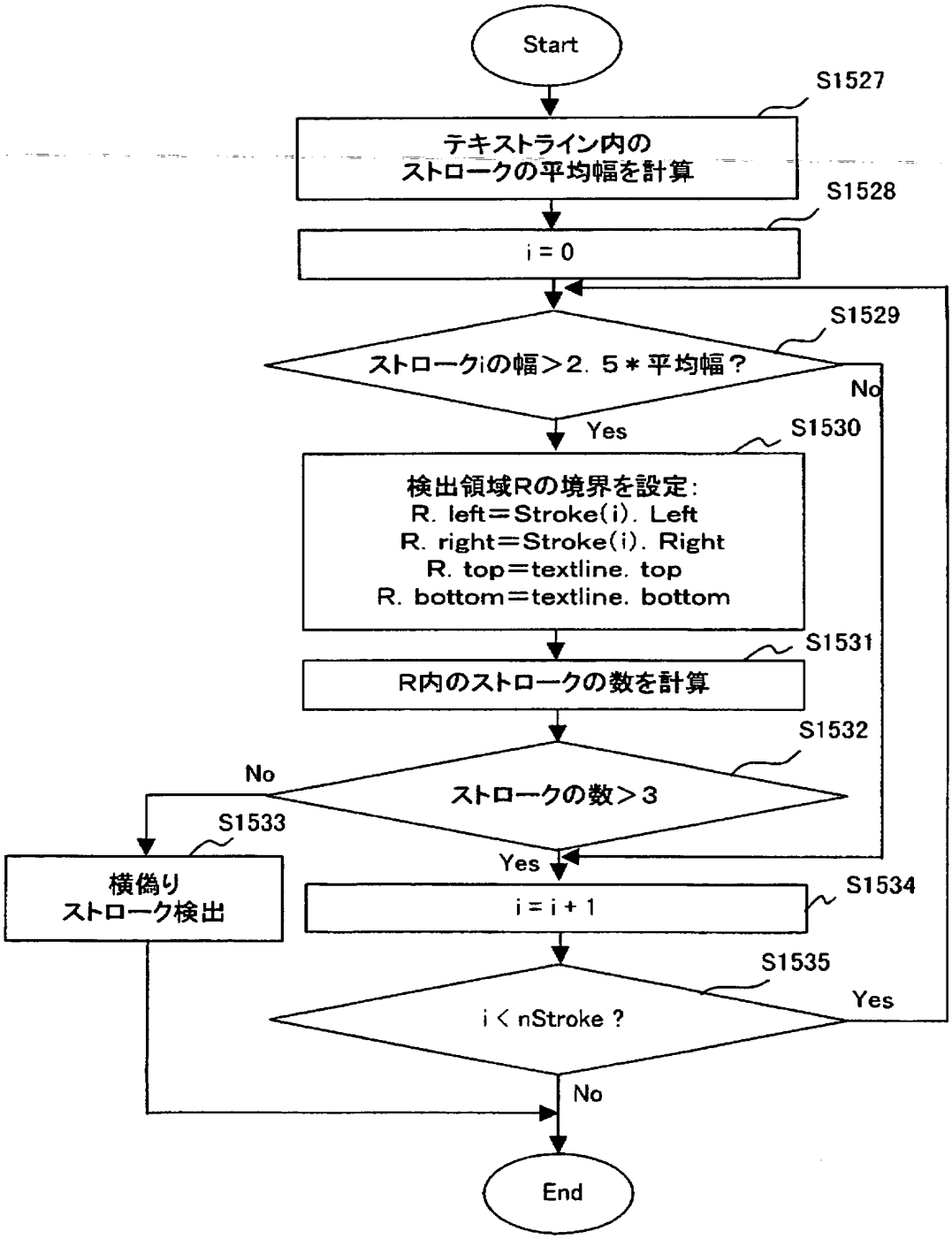
【図５８】

マルチテキストライン検出のフローチャート



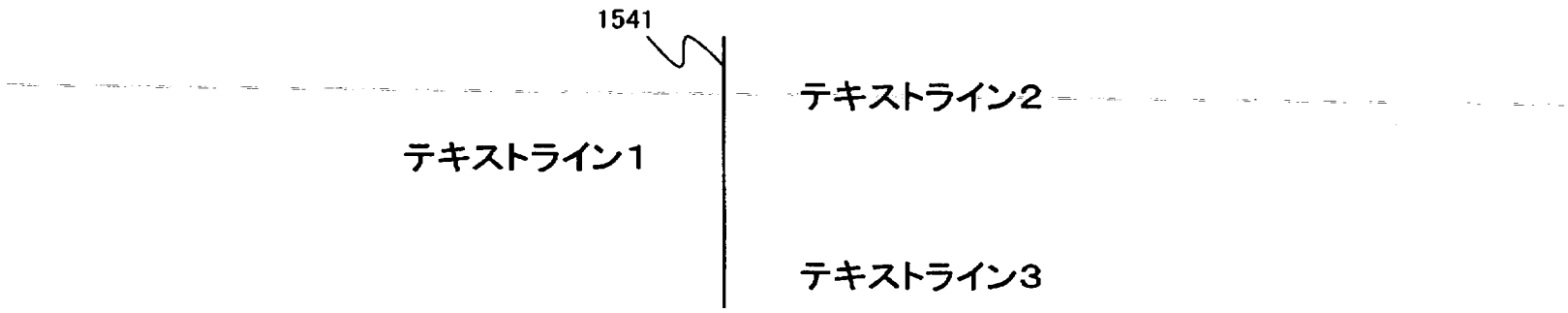
【図５９】

横偽りストローク検出部の動作フローチャート



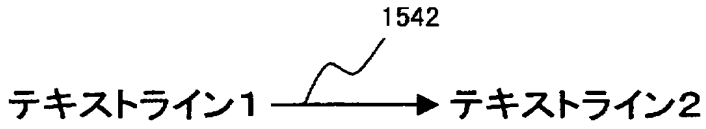
【図６０】

第１の偽りストロークを示す図



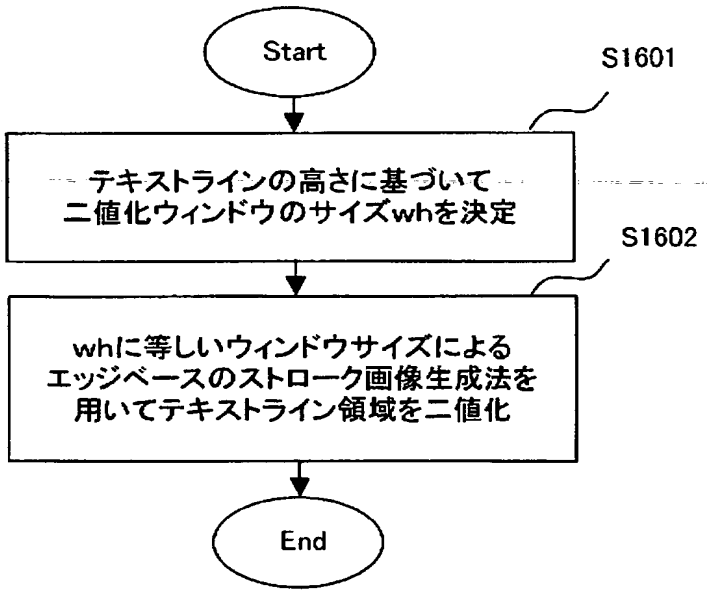
【図６１】

第２の偽りストロークを示す図



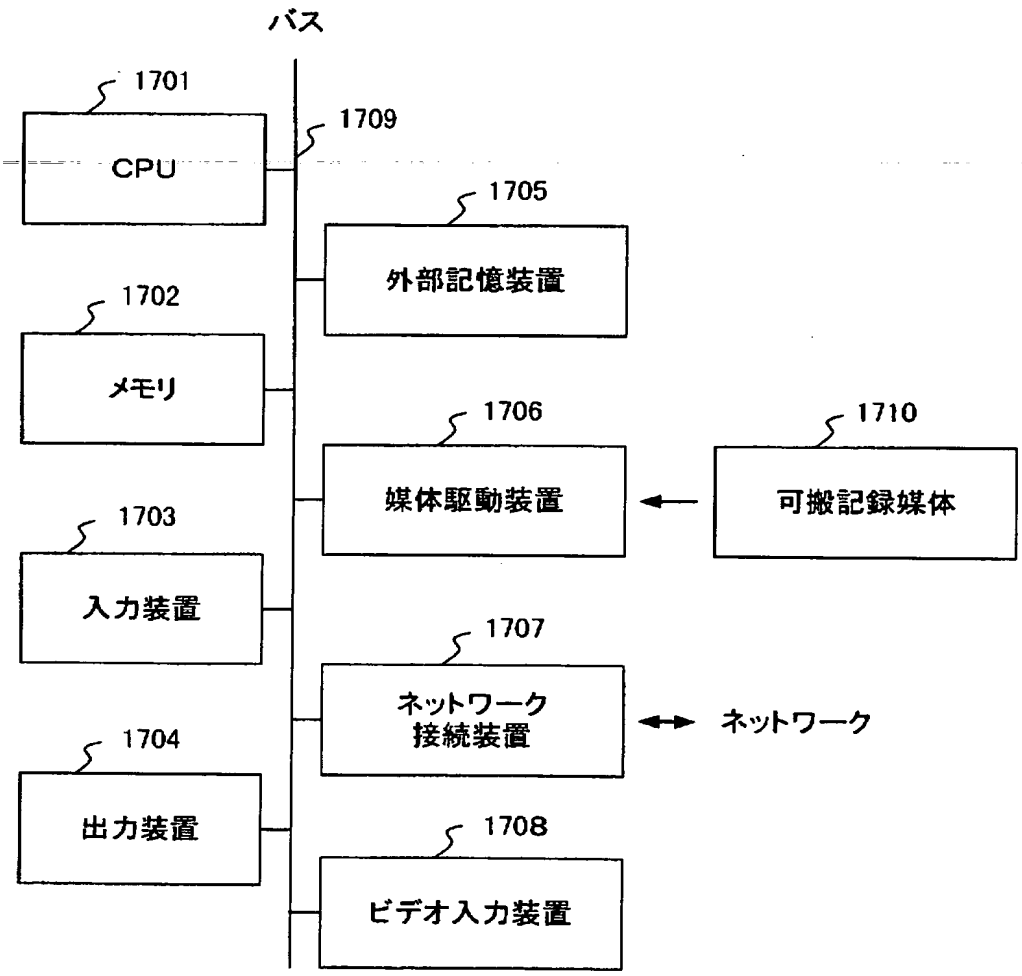
【図６２】

テキストライン二値化部の動作フローチャート



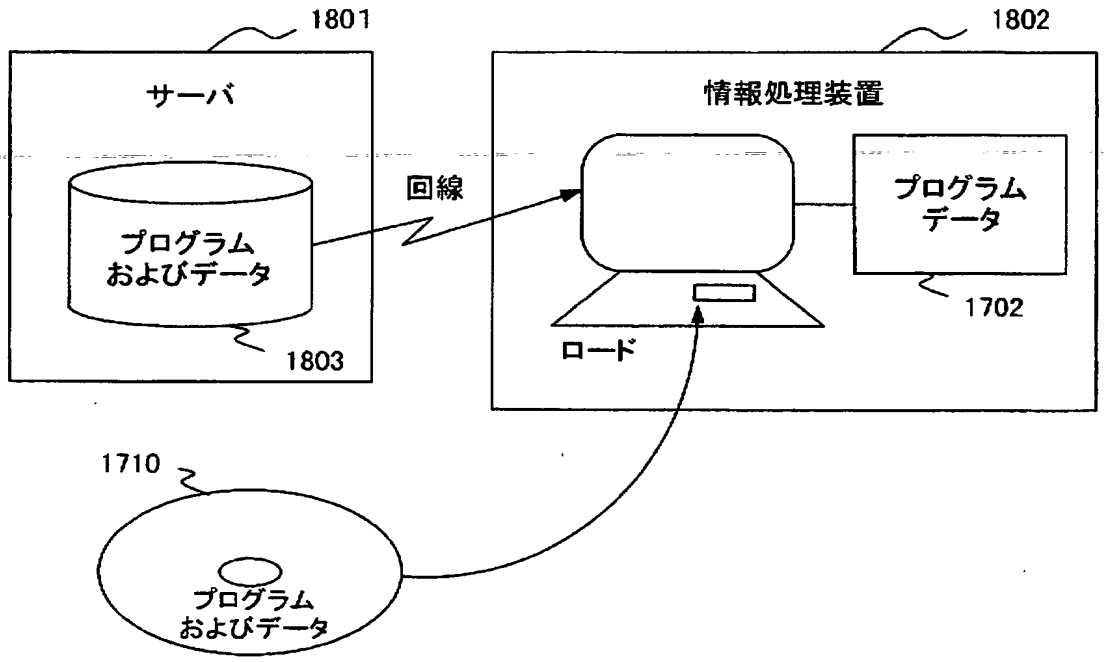
【図６３】

情報処理装置の構成を示す図



【図６４】

記録媒体を示す図



【書類名】　　要約書

【要約】

【課題】　複数のビデオフレームから候補テキストチェンジフレームを高速に選択し、テキストチェンジフレーム内のテキスト領域を正確に検出する。

【解決手段】　与えられたビデオフレームから冗長なフレームや非テキストフレームを除くことにより、テキスト領域を含むビデオフレームを選択し、選択されたフレーム内の偽ストロークを除くことによりテキスト領域を特定し、テキスト領域内のテキストラインを抽出して二値化する。

【選択図】　　図１

特願２００２－３７８５７７

出　願　人　履　歴　情　報

識別番号　　［０００００５２２３］

１．変更年月日　　１９９６年　３月２６日
［変更理由］　　住所変更
住　所　　神奈川県川崎市中原区上小田中４丁目１番１号
氏　名　　富士通株式会社